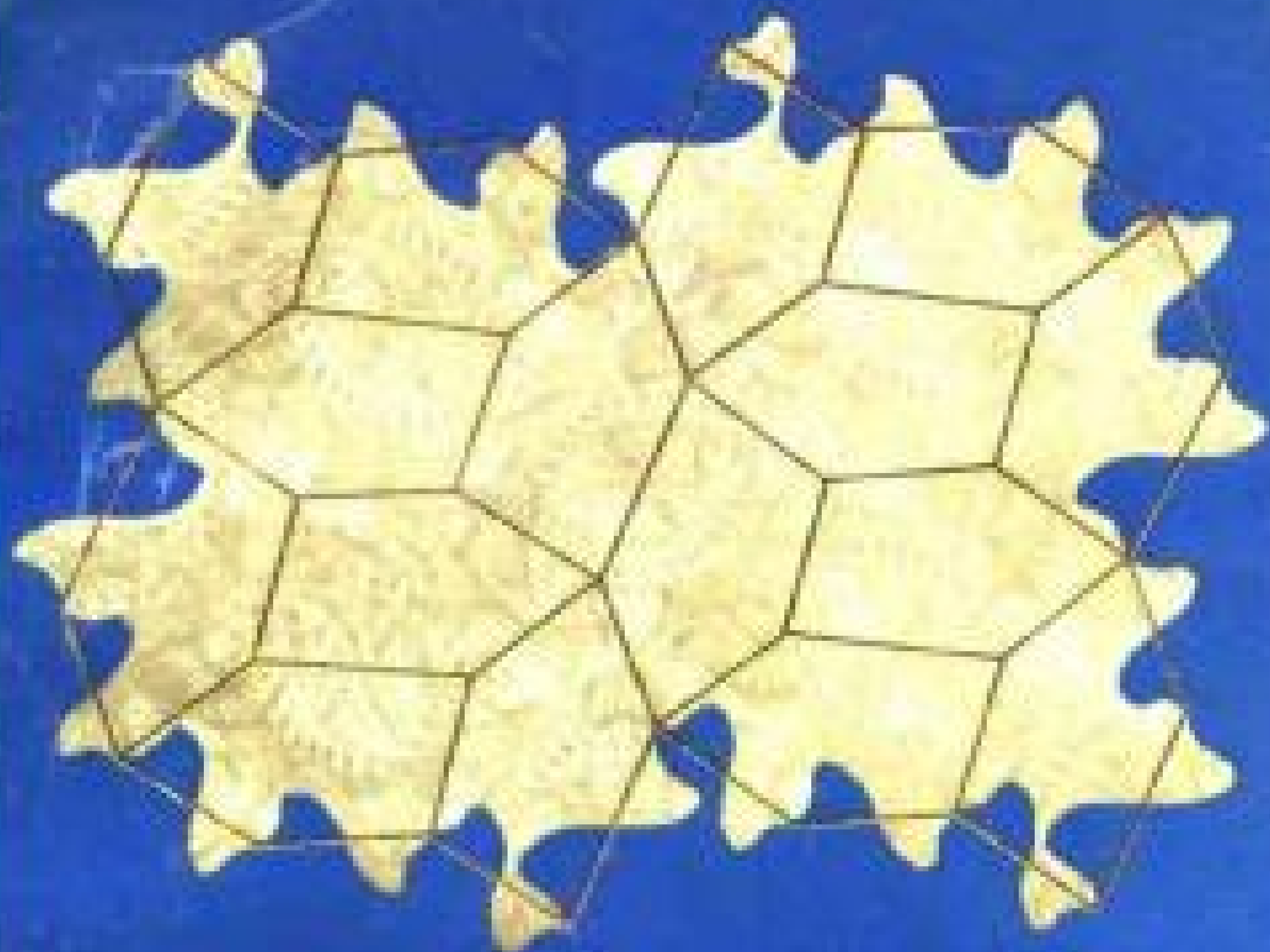


SHU XUE JIADENA

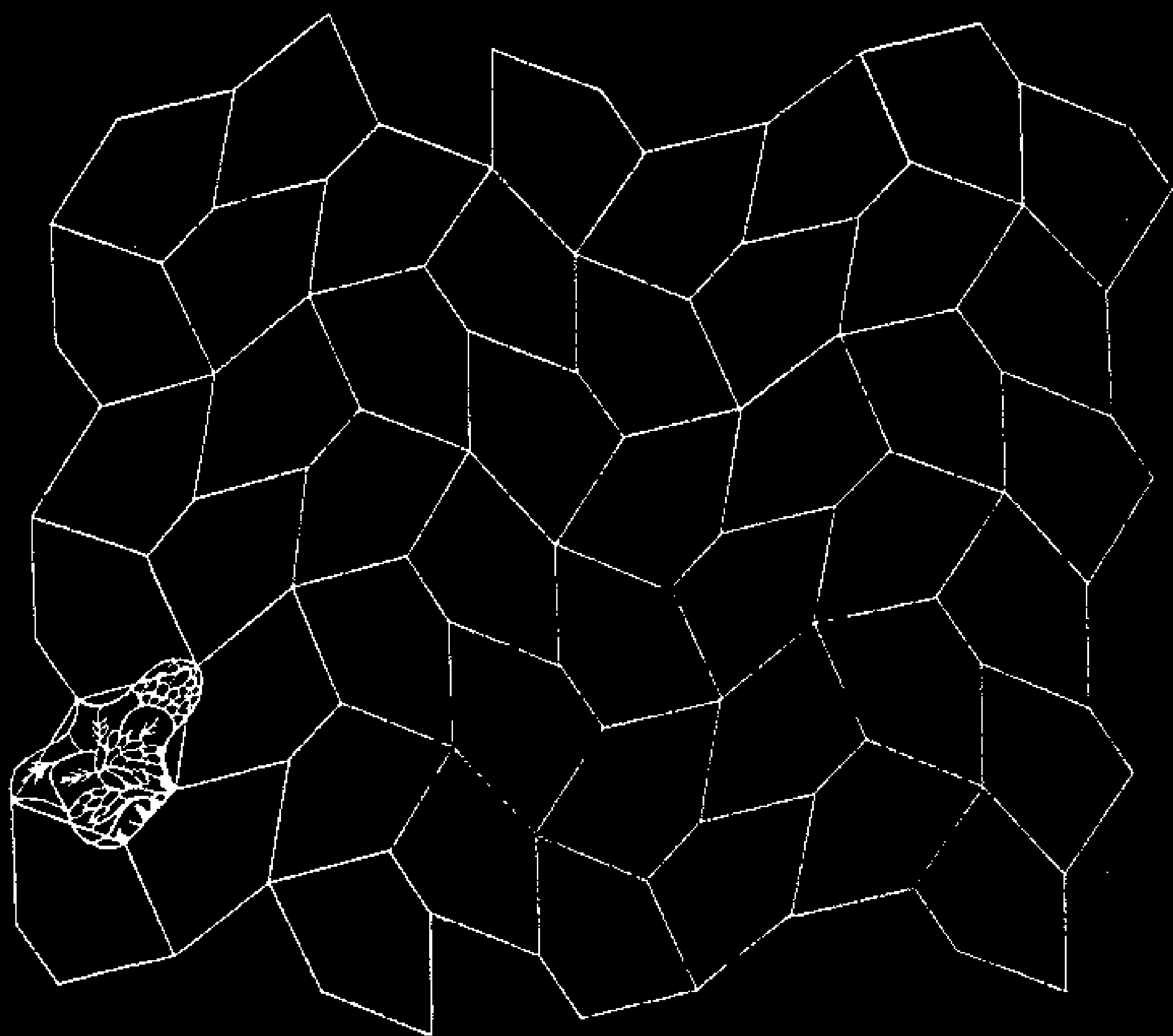
谈祥柏　唐　方译

上海教育出版社

数学加德纳

数学加德纳

谈祥柏　唐　方译 ● 上海教育出版社



The Mathematical Gardner

edited by David A. Klarner

Wadsworth International, 1981

（沪）新登字 107 号

数 学 加 德 纳

[美]戴维·A·克拉纳(David A. Klarner) 编

谈祥柏、唐方 译

上海教育出版社出版发行

（上海永福路 123 号）

各地新华书店经销　　上海群众印刷厂照排、印刷

开本 850×1156 1.32　印张 14.75　插页 6　字数 386,000

1992 年 6 月第 1 版　1992 年 6 月第 1 次印刷

印数 1—2,100 本

ISBN 7-5320-2311 7/G · 2247　　定价:(精)7.90 元

译者序

虽然马丁·加德纳(Martin Gardner，1914 —)从来没有当过教授，但世界各国许多第一流的数学家一听到他的名字，都无不肃然起敬.

马丁·加德纳与数学界的广泛联系已使他成为当今世界上独一无二的人物.在《科学美国人》杂志上每月一篇的专栏文章已经持续二十年以上.这些文章连同他的著作已对现代数学的一些最新成果作了极为出色的通俗介绍.在改善数学的可接受性方面迈出了巨大的步伐.由于他的忘我工作与生花妙笔，已经把一门历来人们认为枯燥乏味的学科，变成了生动活泼有血有肉的"艺术"，吸引了大批青少年投身于数学之门，从而实质上起到了"招兵买马"的作用，为数学立下了汗马功劳.对此，人们也给予他极高的评价，誉之为"数学的传教士"、"数学园丁".甚至以极其崇敬的心情写下了发自内心的颂扬之词：

"在数学这座金碧辉煌、神圣庄严毫不亚于奥林比斯的神庙中，供奉着欧几里得、笛卡儿、牛顿、欧拉、高斯、黎曼、康托等大神，他虽然没有叨陪末座的资格，但是，作为站在庙门口的守护神，那是少不了他的."

马丁·加德纳是一位多才多艺的人物，他在新闻、哲学、文艺等领域都不乏建树，甚至还包括魔术.作为一位"超级魔术家"，他以切身的感受写了不少批判伪科学的文章，令人拍案叫绝.不过，他的主

要业绩，无疑是在数学方面.称之为“数学加德纳”，实在是最切当的.

马丁·加德纳以毕生精力为数学所做的传播工作，其价值无可估量.世界上许多著名数学家在赞赏、理解与支持之余，也觉得理应有所表示.于是，以本书编者戴维·克拉纳(David Klarner)牵头，相约各自撰写一篇自己最拿手的文章，汇集成书，以作为对马丁·加德纳先生65岁寿辰的献礼.此项工作只是在较小的圈子内进行，而并没有大事张扬.因为，马丁·加德纳的崇拜者遍及全球，如果公开征文，那么应征文章必将如雪片飞来，势不可挡.

虽然如此，本书的作者中却拥有为数众多的当代第一流数学家.例如著名的计算机科学大师唐纳德·E·克努特(Donald E. Knuth)(他一人就写了两篇文章)，图论大师贝尔热(Claude Berge)，加拿大多伦多大学教授，伟大的几何学家考克塞特(H. S. M. Coxeter)以及由于发明无法破译的数论“公开密码”而建立了特殊功勋的、麻省理工学院(M. I. T.)的三位专家.荟萃如此众多的专家于一书，在数学出版物中是很少见的.因此本书问世以后，引起了各国学术界的极大重视，新闻媒介也纷纷作了报道.

在众多书评中，国外许多有识之士几乎一致指出：数学是一种看不见的文明，它的触角几乎涉及一切领域(包括文学、艺术、法律、语言等历来认为“与数学风马牛不相及”的学科).在未来的信息化社会(也叫“后工业化社会”)中，传统教学方式必将彻底变革.趣味数学——大脑体操将要发挥它越来越大的作用.也许，21世纪的人类，无论男女老幼，每天都需要做一些智力游戏，就像今天人们喝咖啡、散步、跳舞、上卡拉OK一样.再说，有些尖端的边缘科学，如图象识别、人工智能、混沌——非线性理论等，其本身就是趣味横溢的.

人们往往有一种牢不可破的偏见，认为今天数学的发展非常迅速，业余爱好者们除了欣赏其成果之外，已无用武之地.其实，这种看法有其一定片面性.在本书中有一篇文章专门讲美国的一位家庭妇女玛乔莉·赖斯(Marjorie Rice)的事迹，看了之后着实令人感动.这位子女众多、成天围着锅台转、又未受过正规大学教育的妇女，仍然

在五边形镶嵌问题上作出了前人所没有的发现，足见“科学宝座，人人得而问津，并不是某些人的‘禁区’或‘世袭领地’”.

中国血统的著名数理逻辑学家王浩(H. Wang)业已证明非周期性的铺砌与计算机科学、逻辑学都有紧密联系，即可以决定一个算法以判明给定的一组原始砌块能否铺砌平面. 无独有偶，当代杰出的代数学家，英国剑桥大学著名教授，广义逆矩阵理论的奠基人罗杰·彭罗斯(Roger Penrose)对铺砌问题也表现了巨大的兴趣，他先后发现了风筝、标枪等趣味横溢的构形，其理论非常深刻，而且在结晶学等领域中，有着潜在的应用价值.

考克塞特教授是跨越世纪的数学名著《数学拾零》(Mathematical Recreations and Essays，由剑桥大学三一学院的露斯鲍尔(W. W. Rouseball)于十九世纪中叶写出)的最近一次修订者，他在收入本书的“天使与魔鬼”这一篇文章中，用精炼的有限群论原理对脍炙人口的荷兰大画家埃歇尔(Escher)的一些精彩作品作了透彻的阐明. 看他的文章，无异是一种智慧的享受. 顺便提一下，他的另一篇论文“在纯粹与应用数学研究中美学所起的作用”(The role of aesthetics in pure and applied mathematical research)迄今仍然不失为里程碑式的开创性论文. 人们可以预期，在下一世纪，人们为了填平数学与艺术的鸿沟，它将会起越来越大的作用.

本书的精采篇章很多，我们不可能、也不打算一一介绍. “如人饮水，冷暖自知”，还是请读者自己去鉴赏吧.

常言道：“金无足赤，人无完人.”本书也有许多不足之处. 它给阅者的第一个印象是，全书缺乏统一的体例，听任各位专家自行其是，因而造成笔调不一，风格不一. 有的篇章读起来相当轻松，妙趣横生，有的却行文晦涩，非常吃力. 有的甚至因为省略过多，缺少中间环节而令人感到莫明其妙.

本书的扫描面虽然很广，但对马丁·加德纳的笔触所涉之领域，仍不过是巨厦中之一角. 他所写的一些精彩篇章并未能涉及，特别是与东方文明有关的部分，例如纵横图、易经、折纸、六子联芳以及号称

"中国一绝"的拓扑游戏等内容，这实在是令人遗憾的.

翻译本书的困难很多，主要有下列几点：一是许多作者对数学科普工作并不在行，有些人的母语并非英语，他们所写的文章本身就不是很规范化的；二是某些篇章列有大量文献，因而正文就写得十分简略，而这些文献很难全部看到；三是我国不像日本那样，拥有一支力量雄厚的趣味数学梯队，许多事情只能孤军奋战，缺少相互切磋的机会. 因此，译者虽然付出了艰巨劳动，熬过了许多不眠之夜，但是，疏漏不当之处，肯定在所难免. 深望海内外高明之士，有以教之.

谈 祥 柏

1990 年 9 月写于春申江上.



本书的一些文章都是献给马丁·加德纳的，他是世界上最伟大的数学讲解员与普及大师.我们的文章仅仅局限于这一狭隘领域，然而加德纳的兴趣与成就却涉及许多学科.因此，我们把这本书取名为《数学加德纳》，并且热切地盼望另外的一些书，例如《魔术加德纳》、《文学加德纳》、《哲学加德纳》以及《科学加德纳》等也能相继问世.我们所取的书名自然也是一个双关语，因为马丁·加德纳对数学公众的关系正如园丁同一个美丽花圃的关系[1].这本书的一批作者只不过是一大群数学家中的一小部分，他们的工作受到马丁·加德纳在《科学美国人》上逐月发表的“数学游戏”专栏文章的启发.马丁不单单是数学文章的作者，他还同读者们广泛联系，经常交换各种问题与信息，激励创造活动.台前台后，他都是有力人物.

两位先生在编辑本书的过程中特别有贡献.罗纳德·格兰汉姆(Ronald Graham)与唐纳德·克努特(Donald Knuth)不仅自己撰写了专文，还从其他作者那里征集稿件.编书计划是秘而不宣的，虽然矩阵博士[2]看来已搞到了一些手稿，打算偷偷摸摸地出海盗版.由于出

[1] 译者注：Gardner这个姓与园丁(gardener)仅差一个字母，且发音基本相同.

[2] 译者注：欧文·约书亚·矩阵博士是马丁·加德纳笔下虚构的一位“科学算命”者.加德纳著有《不可思议的矩阵博士》一书，是一本风格独特、学理深奥、尖锐抨击与嘲讽“现代占星术”的巨著.

书计划是悄悄地进行的．作者们都是通过私人接触来选定，而不是公开征文的结果．现在我必须向那些愿意撰稿的人表示歉意．公开征文的呼吁不仅会破坏人们的新奇感，而且必将导致大批稿件雪片般地飞来，势不可挡．

许多人帮助我克服了编辑本书时所遇到的种种棘手问题，作者谨在此表示谢意，其中特别是系主任霍夫曼(Hoffman)先生，在他审阅的文章中编入了他的许多观点，他还鼓励我抓紧完成本书的出版计划，尽管它要占用我大量研究时间．我还得感谢我妻的帮助与耐心．卡拉・林恩(Kara Lynn)在认真准备折迭式中心彩图时扮演了矩阵博士的那位漂亮的欧亚混血种女儿艾娃的角色❶，可惜由于出版商的坚持而只好割爱．在出版这一类书时，显然有大批来往信件，在这方面我的秘书伊丽莎白・牛顿(Elizabeth Newton)的工作确是无可估量的．最后，我要感谢苏珊・格兰汉姆(Susan Graham)以及Prindle，Weber and Schmidt出版公司有关方面的真诚合作，其中，提洛姆・舒立弗(Therom Shreve)特别应当提上一笔．

戴维・A・克拉纳

❶ 译者注：请参看上海科普出版社1990年7月出版的《不可思议的矩阵博士》一书，出版时改名为《"科学算命"之谜》．

目　录

博弈游戏

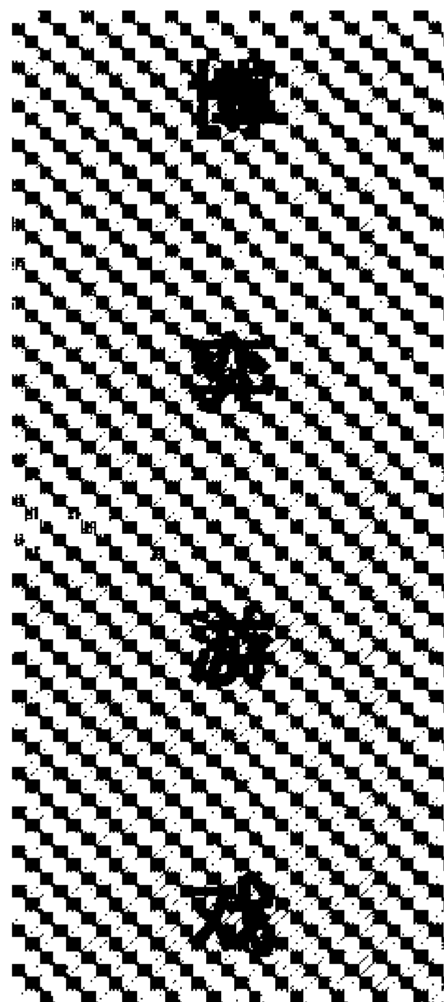
几　何

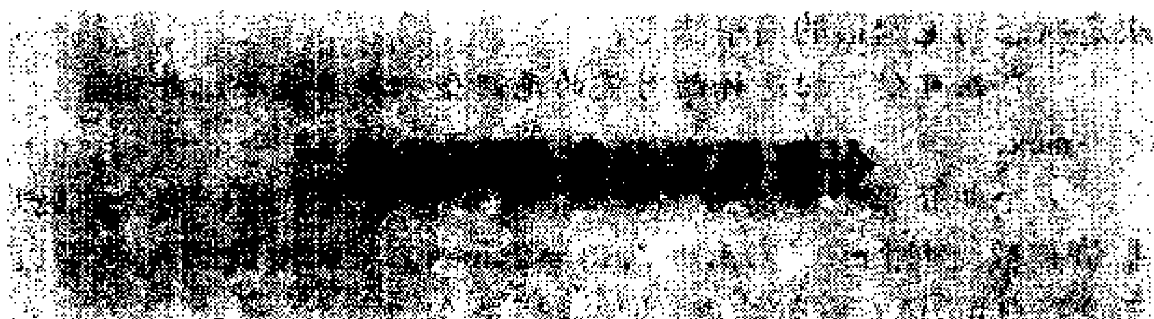
二 维 铺 砌

三 维 铺 砌

文字游戏与动脑筋趣题

数与编码理论


围弈游戏



● 卡尔盖莱大学

〔英〕理查德·K·盖伊(Richard K. Guy)

二柱滚球游戏供两人游玩，其道具是一些小木柱，排列的方法有如图 1 所示. 图上共有八列，每列有 1 或 2 根木柱. 各列间隔得当，以致投球手可以击中任何一列或相邻的两列. 如果含有 2 根木柱的列被击中，则该两柱同时倒下. 在掷一次后，倒下去的木柱在轮到对手掷的时候不再竖起来. 最后 1 根可击的木柱倒下时，本游戏宣告结束，此时的击球手就是赢家. 游戏规则规定：每次击球，必须至少使 2 根木柱倒下去，不准你去击只有 1 根木柱的 1 个列. 因此，这游戏完全有可能在一些孤立的单柱列仍然存在的情况下宣告终止. 例如，在图 1 中，你可以只是把 d 列击倒，但不能仅仅只击中 b、c、e 或 h 列，除非你同时还击中了其他相邻的列. 在 d 列倒下去之后，对方不可能同时击倒 c、e 列，因为它们是不相邻的.

二柱滚球游戏被认为是一种无偏袒的博弈，因为在任何状态下，可供选择的策略集合对两位局中人来说，都完全一样. 反之，国际象棋则是一种有偏袒的博弈. 因为，在任何状态，黑方与白方可选择的策略集合并不相同. 最后一位游戏者被认为是赢家的无偏博弈理论并未为人广泛通晓，虽说它理应如此. 这一理论是 Sprague〔21〕与 Grundy〔12〕以及其他人先后互相独立地发现的. 他们发现：在任何无偏博弈中，每个状态都具有一个*尼姆值*，这就是说，它等

价于一个尼姆堆，即尼姆游戏〔4，2，15〕中的一堆豆子．有一个决定状态的尼姆值的简单法则：

在可供自由选择的对策的尼姆值中取其最小排外值 mex.

一个非负整数集的最小排外值 mex 是不属于这一集合的最小非负整数．例如 mex{5,3,0,7,1}=2，mex∅=0，由此可见终局（已无选择自由，游戏宣告结束）的尼姆值是 0.

尼姆值也称为 Sprague-Grundy 函数，它的重要性来自下列事实：无偏博弈的一切状态构成一个阿贝尔加法群．事实上，凡是最后下着算赢的一切博弈（其中也包括有偏袒的）都是这样，不过，Sprague-Grundy 理论只能适用于无偏博弈的那个子群.

两个或两个以上状态（不一定要求它们属于同一种博弈游戏）的和（也称为析取）可按如下方式进行：

轮到他走的局中人可任意选择一个子博弈，并在其中走出合乎规则的一步.

当每一个子博弈都结束时，复合博弈宣告结束，最后能下着者算是赢家．容易看出，这种加法能满足交换律与结合律.

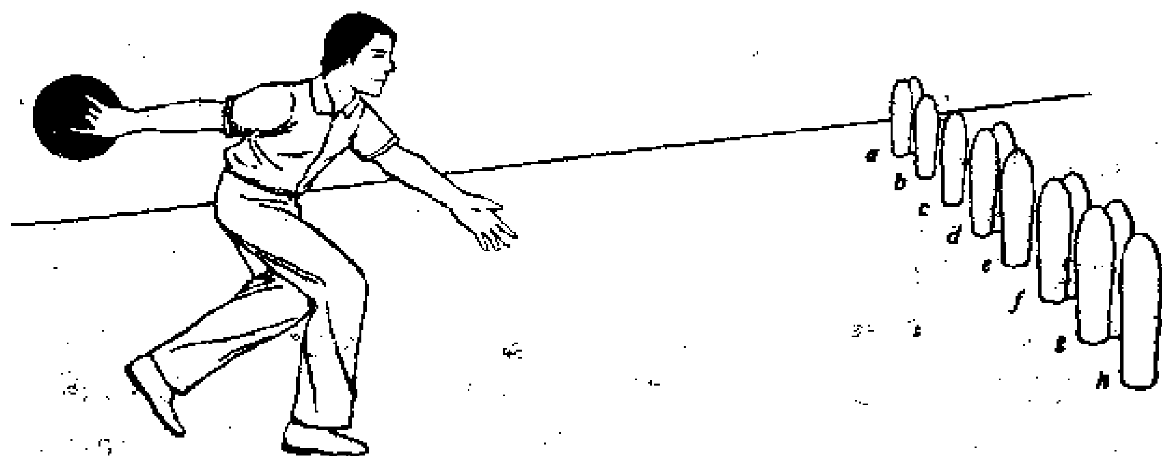


图 1

在二柱滚球游戏中准备一掷

群的幺元(单位元)显然是终局,而任意状态的负状态则仍然是那个状态,不过是由对方先下着(在无偏博弈中任一状态的负状态就是其本身).绝大多数人在无意中发现了所谓"亦步亦趋"原则得以适用的博弈游戏.这时你可以采用一种后发制人的对称策略,即模仿对方的一招一式,依样画葫芦,直至你取得胜利.这种加法群不仅在数学上非常美,而且在实用上也是重要的,因为许多博弈游戏在正常进行的过程中可以分解成一些子博弈的和.例如,二柱滚球游戏相当典型的一着是把一行分裂为两短行,于是,下一步就必须在两行之一中采取行动.

最后下着算作赢家的无偏博弈之Sprague-Grundy理论的主要结果可归纳为下列定理:

两个博弈之和的尼姆值是它们的尼姆值的尼姆和.

要想求出两个非负整数的尼姆和,只要对它们施行没有进位的二进制加法.这就是Bouton[4]在其分析尼姆游戏(也可参看[2]与[15])的原始论文中所用的运算.既然我们有了Sprague-Grundy理论,尼姆即可视为一切无偏博弈的原型:典型的尼姆状态是尼姆游戏的析取和,而在每个子博弈中,人们都在和一堆东西打交道.

二柱滚球游戏由埃尔文·伯勒坎普(Elwyn Berlekamp)首先发现,在他对众所周知的纸笔游戏点与盒(也叫点与方)[10]所作的巧妙解析[见文献3的第16章]一文中可以看到.它也包罗了作为其特例的开勒司(Kayles)游戏[8,19,9]以及道森氏开勒司(Dawson's Kayles)游戏[6,7],我们将在下文加以描述,而且其解析也是已知的.事实上,Guy与Smith[14]❶已经研究了一大族"取子与分开"游戏,玩这些游戏的道具是排成几行或集成几堆的豆子.它们也可以称

❶ 译者注:这是一篇非常重要的论文,原文载于《*Proceedings of Cambridge Philosophical Society*》杂志第52卷514—526页.译者可提供影印件.建议有兴趣的读者首先要阅读一下.

为八进码游戏，因为游戏规则可通过八进码来加以描述：

$$d_0.\ d_1d_2d_3\cdots\cdots$$

这里 $d_0=0$ 或 4（把一行或一堆豆子分为两行或两堆，不准有空集，也不取走任何一粒豆子），$0\leqslant d_r\leqslant 7$，对 $r\geqslant 1$；各位数码的意义见表 1. 例如，被丢得涅（Dudeney）称为 Kayles〔8〕，洛伊德（Loyd）称为 Rip Van Winkle❶ 的游戏〔19，9〕，用这种八进码表示时就是 0.77. 事实上，它是二柱滚球游戏的特例，即当每一列都恰有两根木柱. 于是，游戏规则可简明地概括为：拿走 1 列或 2 个相邻的列.

T·R·道森所提出的一个问题促成了八进码游戏的数学分析，他是"侏儒象棋"专家〔6，7〕，该问题称为道森氏象棋. 下这种棋的棋盘只有 3 行 n 列（见图 2❷）白方与黑方的兵卒分别位于第 1 和第 3 行，吃子是强制性的（可吃的子一定要吃，不能不吃），棋赛的性质是一种比较特殊的"输棋"，就是说，能走最后一步的人算输. 凡是懂得国际象棋的兵卒走法以及吃子规则的人马上就会看出，在双方交替吃子之后，一对兵卒就陷入相互顶牛的状态而丝毫不能动弹❸. 由此可见，道森氏象棋实际上与一行豆子游戏等价. 在该游戏中，可以取走任何一颗豆子，但必须把它的紧邻豆子同时取走.（如豆子在中间，则必须把它的左、右紧邻豆子同时取走；如豆子在边上，则必须把紧挨它的一颗豆子同时取走；如为孤立的豆子，则不能取.）你可以用八进码进行对照，就会知道这种游戏的代码是 0.137.

---

❶ 译者注：Rip Van Winkle，人名，他一觉醒来，世事已如沧海桑田，发生巨变. 其情节类似于我国《烂柯神机》、刘、阮入天台采药等故事. 见《Sam Loyd 游戏数学百科全书》，原书有图.

❷ 译者注：实际上是 3 行 8 列，此处 $n=8$.

❸ 译者注：由于棋盘上一共只有 3 行，所以"吃过路兵"等复杂情况不会发生. 如第一列的白卒向前进一步，则第二列的黑兵便斜走一格把它吃掉，此时第二列的白卒马上又把黑兵吃掉，于是第一列的白卒与黑兵即陷入顶牛状态. 其余情况请读者自己考虑.

$d_r=$	在几行豆子上所做的游戏	在几堆豆子上所做的游戏
0	不存在合乎规则的行动来取走 r 颗豆子.	
$1=2^0$	如 r 颗豆子组成一个整行，则它们都可以取走.	一整堆 r 颗豆子都可以完全取走.
$2=2^1$	从一个较长行的任何一头都可取走 r 颗豆子.	可以从较多豆子的一堆中取走 r 颗豆子，使得该堆中还有豆子留下.
$3=2^1+2^0$	对以下两种情况中的任一种情况，都可以取走 r 颗豆子， （留下 0 行或 1 行）.	（留下 0 堆或 1 堆）.
$4=2^2$	在较长一行的内部可以取走 r 颗连在一起的豆子，使得留下的两行都有豆子.	在含有 $r+2$ 颗豆子或更多豆子的一堆中可以取出 r 颗豆子，使留下的两堆各有豆子.
$5=2^2+2^0$	在一行豆子中可以取走 r 颗连续的豆子，如果留下的是 0 行或 2 行的话.	有 r 颗豆子的一堆可以完全取光，或者从豆子数 $\geqslant r+2$ 的一堆中取走 r 颗，使剩下的两堆都不空.
$6=2^2+2^1$	可以从较长的一行中取走 r 颗连续的豆子，使留下的 1 行或 2 行都不是空的.	可以从豆子数较多的一堆中取出 r 颗豆子，使留下的 1 堆或 2 堆中都有豆子.
$7=2^2+2^1+2^0$	在下列情况下可以取走 r 颗豆子， （留下 0，1 或 2 行）.	（留下 0，1 或 2 堆）.

表 1

八进数码 d_r 的意义.

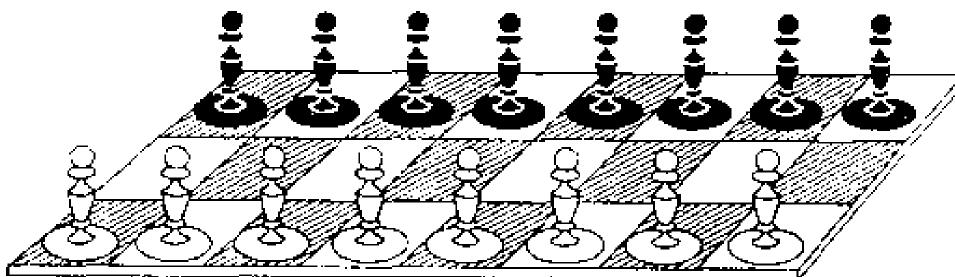


图 2

双方准备下道森氏象棋.

正如道森当时所建议的那样，这种棋戏的胜负规定是取的异常形式，也就是说最后能走子者算是输家.异常形式的博弈较之正常形式（最后能行动者算是赢家）的博弈分析起来要远为困难得多.由于异常形式的尼姆值只要在接近收尾时略为修改一下策略就行，这就给人们造成一种错觉，似乎对别的无偏博弈所取的对策也可以类似地稍加修改.然而，对绝大多数游戏来说，这种估计是不正确的.（请参看 Grundy 与 Smith 的文章〔13〕或 Conway 一文〔5 的第 12 章〕，他给出了异常形式的道森氏象棋的最初几种状态的分析（在八进码代号0.4游戏以及 Kayles 中〔第 145 页〕，更详细的分析可参看文献〔2〕的第 16 章.）

不难证明〔14〕，代码为 0.137，0.07 与 0.4 的几种游戏是相互紧密联系着的.我们把 0.07 这种形式的游戏称为道森氏开勒司，它可以用排成一行的豆子来做游戏；每一步定义为取走两颗相邻豆子.由此可见，它是二柱滚球游戏的一个特例，这时每列只有 1 根木柱.开勒司与道森氏开勒司游戏，在用排成一行的 n 颗豆子来玩时，其尼姆值已被人发现〔14〕具有周期性，除了对很小的 n 值有点不规则以外，上述两种游戏尼姆值的周期分别是 12 与 34.

对二柱滚球游戏作出完整的分析显然是不现实的，因为它的状态非常之多.对 n 列来说，究竟有多少种本质上不相同的状态呢？由于正好有两种列，简单的答案便是 2^n 个状态，不过我们并不需要研究所有这一切状态，因为伯勒坎普业已指出状态之间的一些等价关系，你也容易验证其正确性：

1. $0*ijk\cdots=*ijk\cdots=00ijk\cdots$,
2. $\cdots ijk*0*lmn\cdots=\cdots ijk*+*lmn\cdots$,
3. $\cdots ijk*00*lmn\cdots=\cdots ijk*\ *\ *lmn\cdots$,

这里 0 表示只有 1 根木柱的列（此种情况，双方都不能击倒它，因此它可以维持到终局）；* 表示有 2 根木柱的列（任意一方都可以

击倒它，因此它相当于尼姆值为 1 的堆）. 游戏之星〔见 5 的第 72 页〕与字母则表示两种列中的任意一种. 在第 2 式右边，加法运算符号即表示我们已描述过的析取和.

于是，我们只需分析那些两端有一个星号的二柱滚球游戏状态（例如 1 式）以及不出现 0（只有 1 根木柱的列或者至少有三堆才出现 0）的情况（如 2 或 3 式）. 具有此种性质的二进位序列已被 Austin 与 Guy〔1〕彻底列举过，仅不过他们用 0 与 1 来代替我们的 * 与 0 而已. 若使用他们的记号，这类二柱滚球游戏的有关数 t_n 便是 $a_{n-2}^{(3)}$，之所以要减去 2，那是因为在行的两端有两个星号之故. 数 t_n 满足下列递归关系式

$$t_n = 2t_{n-1} - t_{n-2} + t_{n-4}.$$

事实上

$$t_n = \frac{1}{2}f_n + \frac{1}{\sqrt{3}}\sin\left(\frac{n\pi}{3}\right),$$

这里的 f_n 是斐波那契数.

4. $$f_n = \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right\}.$$

然而，我们没有必要去分析与那些已分析过的状态有反射关系的状态，所以，我们在下面将探讨对称状态数 S_n.

一个对称状态的中心是四种类型 A、B、C、D 之一，已表示于图 3 的左方. 若 n 为奇数，则记号？代表 0 或 *. 若 n 为偶数，则用一对相当的记号来代替中心的记号. 中心记号（若 n 为奇数时）可以被下面的记号取代

(a) * * *， (b) 0 * 0， 或 (c) 0 0 0

用以产生多出两列的对称状态，除非 (a) 不一定用于 B 与 C，(b) 不一定用于 A 或 B.

若 n 为偶数，则将中间的一对记号用下面的记号予以取代

(a) * * * *， (b) 0 * * 0， 或 (c) 0 0 0 0.

设以 A_n 表示属于类型 A 的对称 n 列状态数，B_n，C_n，D_n 分别表示类型 B、C、D 的，则

$$A_n = A_{n-2} + D_{n-2},$$

$$B_n = A_{n-2},$$

$$C_n = B_{n-2},$$

$$D_n = C_{n-2} + D_{n-2}.$$

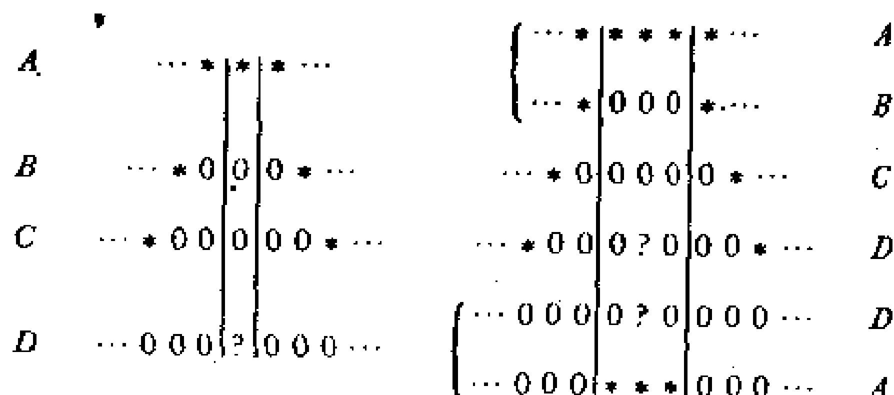


图 3

一个对称状态的四种中心类型.

考虑到 D 的两种歧义，在计算总数时插入一个系数 2，

$$\begin{aligned} s_n &= A_n + B_n + C_n + 2D_n \\ &= (A_{n-2} + B_{n-2} + C_{n-2} + 2D_{n-2}) + (A_{n-2} + C_{n-2} + D_{n-2}) \\ &= (A_{n-2} + B_{n-2} + C_{n-2} + 2D_{n-2}) + (A_{n-4} + B_{n-4} + C_{n-4} + 2D_{n-4}) \end{aligned}$$

于是可知，s_n 满足以下递推关系

5. $$s_n = s_{n-2} + s_{n-4}.$$

其值为

$$s_n = f_{\lfloor (n+1)/2 \rfloor}$$

这里$\lfloor\ \rfloor$为底价函数（不大于自变量的最大整数），而 f 则是斐波那契数，见公式 4.

因此，不对称的二柱滚球游戏的状态数，在不把反射状态看作不同的情况下，可计算如下：

$$u_n=\frac{1}{2}(t_n-s_n)=\frac{1}{4}f_n-\frac{1}{2}f_{\lfloor (n+1)/2\rfloor}+\frac{1}{2\sqrt{3}}\sin\frac{n\pi}{3},$$

而不计反射的总数是

$$v_n=\frac{1}{2}(t_n+s_n)=\frac{1}{4}f_n+\frac{1}{2}f_{\lfloor (n+1)/2\rfloor}+\frac{1}{2\sqrt{3}}\sin\frac{n\pi}{3}.$$

更一般的情形

$$t_n^{(k)}=a_{n-2}^{(k)}$$

在文献〔1〕中进行了充分讨论，该处记号 0 连续出现的长度至少是 k. 这里我们将把上述数学解析加以推广，以获得对一般的 k 的相应序列 $s_n^{(k)}, u_n^{(k)}$ 以及 $v_n^{(k)}$. 对 $k\geqslant 1$，公式一般为真，但对 $k=1$，没有蕴含什么限制条件(除却规定两端均要有 * 记号)，易于看出，对 $n\geqslant 2$(以及 $k=1$)，下面这些式子成立：

$$t_n=2^{n-2}, s_n=2^{\lfloor (n-1)/2\rfloor}, u_n=2^{n-3}-2^{\lfloor (n-3)/2\rfloor}, v_n=2^{n-3}+2^{\lfloor (n-3)/2\rfloor}.$$

从现在起，我们将省略上标(k).

首先，我们利用关系式(见〔1〕)，

6. $\quad t_m=2t_{m-1}-t_{m-2}+t_{m-k-1},$

故有

$$\begin{aligned} t_m-t_{m-1} &= t_{m-1}-t_{m-2}+t_{m-k-1} \\ &= t_{m-2}-t_{m-3}+t_{m-k-1}+t_{m-k-2} \\ &\cdots\cdots \\ &= t_{k+1}-t_k+t_{m-k-1}+t_{m-k-2}+\cdots+t_2+t_1. \end{aligned}$$

由于 $t_1=t_2=\cdots=t_k=t_{k+1}=1$，我们有

7. $\quad t_m-t_{m-1}=\sum_{i=1}^{m-k-1}t_i,$

这是一个计算$\{t_m\}$的方便算法. 我们也可就此公式进行求和以获得

8. $\quad t_m=1+\sum_{i=1}^{m-k-1}(m-k-i)t_i.$

在文献〔1〕中并未给出公式 7 与 8.

其次，我们将在更一般的形式下建立起公式 5.

9. $s_n = s_{n-2} + s_{n-k-1}$.

情形 A $k=2l-1$ 为奇数，$n=2m-1$ 或 $2m$.

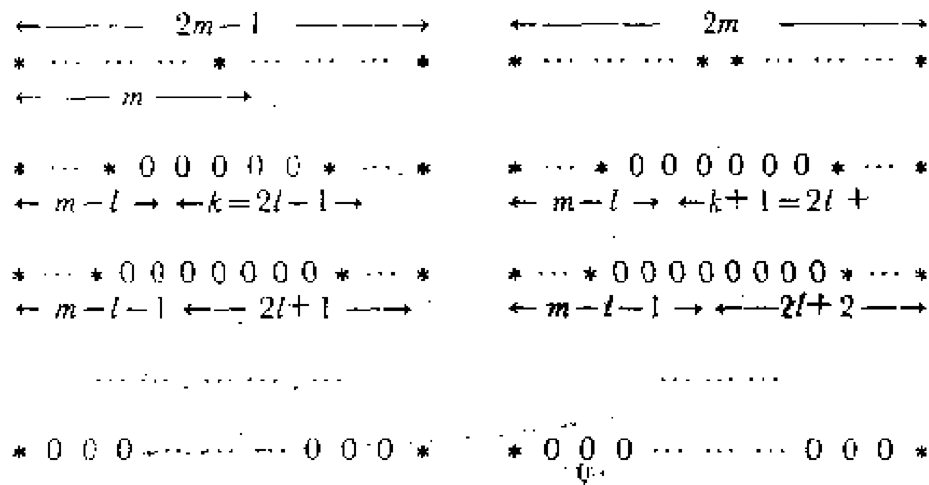


图 4

连续出现 0，其长度≥k 时，二柱滚球游戏的对称状态.

由图 4，我们可以看出，对称状态数是

$$s_n = s_{2m-1} = s_{2m} = t_m + t_{m-l} + t_{m-l-1} + \cdots + t_1$$

$$s_{n-2} = s_{2m-3} = s_{2m-2} = t_{m-1} + t_{m-l-1} + t_{m-l-2} + \cdots + t_1$$

$$s_n - s_{n-2} = t_m - t_{m-1} + t_{m-l}$$

$$= t_{m-l} + \sum_{i=1}^{m-2l} t_i.$$

由 7 式，故得 $s_n - s_{n-2} = s_{2(m-l)-1} = s_{2(m-l)} = s_{n-k-1}$，这便是我们所要的结果.

情形 B $k=2l$（偶数），类似情形 A，但我们必需分别处理 $n=2m$ 与 $n=2m-1$：

$$s_{2m} = t_m + t_{m-l} + t_{m-l-1} + \cdots + t_1,$$

$$s_{2m-1} = t_m + t_{m-l-1} + t_{m-l-2} + \cdots + t_1,$$

$$s_{2m-2} = t_{m-1} + t_{m-l-1} + t_{m-l-2} + \cdots + t_1,$$

$$s_{2m-3} = t_{m-1} + t_{m-l-2} + t_{m-l-3} + \cdots + t_1,$$

$$s_{2m}-s_{2m-2}=t_m-t_{m-1}+t_{m-k},$$

$$s_{2m-1}-s_{2m-3}=t_m-t_{m-1}+t_{m-k-1},$$

利用 7 式，我们在这两种情况下都能得出 9 式，同前面一样.

给出 t_n 与 s_n 的母函数是

$$T(z,k)=\sum_{i=0}^{\infty}t_i^{(k)}z^i=\frac{z(1-z)}{(1-z)^2-z^{k+1}},$$

$$S(z,k)=\sum_{i=0}^{\infty}s_i^{(k)}z^i=\frac{z(1+z)}{1-z^2-z^{k+1}}.$$

利用 7 式与 9 式，再加上

$$u_n=\frac{1}{2}(t_n-s_n),v_n=\frac{1}{2}(t_n+s_n)$$

就能够使我们计算表 2 中的一些值，表中的点号“…”表示在此之前，对应于较小正值 n 的有关数列都是常数.

在这些数列中，出现于 Sloane 手册〔20〕中的只有下面几种：

2 的幂，$t_n^{(1)}=2^{n-2}$,

斐波那契数，$s_n^{(3)}$，以及

序列 # 102，$s_n^{(2)}$.

最后的这个数列在文献〔11〕中出现过，它是作为帕斯卡三角形中位于广义对角线

$$3x+2y=n-1$$

上的各数之和的实例而被提及的. 在文献〔16，17，18〕中，述及因式分解与讨论可除性时也讲到过它. 例如，

若 $n=7m-3,7m-2$ 或 $7m$，则 $s_n^{(2)}$ 是偶数.

能整除 $s_{7m-3}^{(2)}$ 的 2 的最高次幂称为“直尺函数”，即 $2m$ 中 2 的最高次幂.

若 $n=13m-3,13m-2,13m$ 或 $13m+6$，则 3 能整除 $s_n^{(2)}$.

对二柱游戏来说，我们已经有一点离题太远了. 在图 1 中，最好的一击是什么呢? 伯勒坎普的等价方程(即 1 式)告诉我们，h 列可以忽略不计，第 2 式告诉我们，可以移走 e 列而不影响状态. 等价关系

	n	…3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	t_n	…1	2	4	7	12	21	37	65	114	200	351	616	1081	1897	3329	5842	10252
$k=2$	s_n	…1	2	2	3	4	5	7	9	12	16	21	28	37	49	65	86	114
	u_n	…0	0	1	2	4	8	15	28	51	92	165	294	522	924	1632	2878	5069
	v_n	…1	2	3	5	8	13	22	37	63	108	186	322	559	973	1697	2964	5183

20	21	22	23	24	25	26	27	28	29	30
17991	31572	55405	97229	170625	299426	525456	922111	1618192	2839729	4983377
151	200	265	351	465	616	816	1081	1432	1897	2513
8920	15686	27570	48439	85080	149405	262320	460515	898380	1418916	2490432
9071	15886	27835	48790	85545	150021	263136	461596	809812	1420813	2492945

	…4	5	[illegible]	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	…1	2	[illegible]	7	11	17	27	44	72	[illegible]	189	305	493	798	1292	2091	3383	5473
$k=3$	…1	2	2	3	3	5	5	8	8	13	13	21	21	34	34	55	55	89
	…0	0	1	2	4	6	11	18	32	52	88	142	236	382	629	1018	1664	2692
	…1	2	3	5	7	11	16	26	[illegible]	65	101	163	257	416	663	1073	1719	2781

22	23	24	25	26	27	28	29	30	31	32
8355	14328	23184	37513	60697	98209	158905	257114	416020	673135	1089155
89	144	144	233	233	377	377	610	610	987	987
4383	7092	11520	18640	30232	48916	79264	128252	207705	336074	544084
4472	7236	11664	18873	30465	49293	79641	128862	208315	337061	545071

	…5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	…1	2	4	7	11	16	23	34	52	81	126	194	296	450	685	1046	1601	2452	3753
$k=4$	…1	2	2	3	3	4	5	6	8	9	12	14	18	22	27	34	41	52	63
	…0	0	1	2	4	6	9	14	22	36	57	90	139	214	329	506	780	1200	1845
	…1	2	3	5	7	10	14	20	30	45	69	104	157	236	356	540	821	1252	1908

24	25	26	27	28	29	30	31	32	33	34	35
5739	8771	13404	20489	31327	47904	73252	112004	171245	261813	400285	612009
79	97	120	149	183	228	280	348	429	531	657	811
2830	4337	6642	10170	15572	23838	36486	55828	85408	130641	199814	305599
2909	4434	6762	10319	15755	24066	36766	56176	85837	131172	200471	306410

	…6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
	…1	2	4	7	11	16	22	30	42	61	91	137	205	303	443	644	936	1365	1999
$k=5$	…1	2	2	3	3	4	4	6	6	9	9	13	13	19	19	28	28	41	41
	…0	0	1	2	4	6	9	12	18	26	41	62	96	142	212	308	454	662	979
	…1	2	3	5	7	10	13	18	24	35	50	75	109	161	231	336	482	703	1020

25	26	27	28	29	30	31	32	33	34	35	36	37
2936	4316	6340	9300	13625	19949	29209	42785	62701	91917	134758	197548	289547
60	60	88	88	129	129	189	189	277	277	406	406	595
1438	2128	3126	4606	6748	9910	14510	21298	31212	45820	67176	98571	134476
1498	2188	3214	4694	6877	10039	14699	21487	31489	45097	67582	98977	145071

$k=6$	…7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	…1	2	4	7	11	16	22	29	38	51	71	102	149	218	316	452	639	897	1257
	…1	2	2	3	3	4	4	5	6	7	9	10	13	14	18	20	25	29	35
	…0	0	1	2	4	6	9	12	16	22	31	46	68	102	149	216	307	434	611
	…1	2	3	5	7	10	13	17	22	29	40	56	81	116	167	236	332	463	646

26	27	28	29	30	31	32	33	34	35	36	37	38
1766	2493	3536	5031	7165	10196	14484	20538	29085	41168	58282	82561	117036
42	49	60	69	85	98	120	140	169	200	238	285	336
862	1222	1738	2481	3540	5049	7182	10199	14458	20484	29022	41138	58350
904	1271	1798	2550	3625	5147	7302	10339	14627	20684	29260	41423	58686

$k=7$	…8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	…1	2	4	7	11	16	22	29	37	47	61	82	114	162	232	331	467	650	894
	…1	2	2	3	3	4	4	5	5	7	7	10	10	14	14	19	19	26	26
	…0	0	1	2	4	6	9	12	16	20	27	36	52	74	109	155	224	312	434
	…1	2	3	5	7	10	13	17	21	27	34	46	62	88	123	175	243	338	460

27	28	29	30	31	32	33	34	35	36	27	38	39
1220	1660	2262	3096	4261	5893	8175	11351	15747	21803	30121	41535	57210
36	36	50	50	69	69	95	95	131	131	181	181	250
592	812	1106	1523	2096	2912	4040	5628	7808	10836	14970	20677	28480
628	848	1156	1573	2165	2981	4135	5723	7939	10967	15151	20858	28730

	···9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	···1	2	4	7	11	16	22	29	37	46	57	72	94	127	176	247	347	484	667
$k=8$	···1	2	2	3	3	4	4	5	5	6	7	8	10	11	14	15	19	20	25
	···0	0	1	2	4	6	9	12	16	20	25	32	42	58	81	116	164	232	321
	···1	2	3	5	7	10	13	17	21	26	32	40	52	69	95	131	183	252	346

28	29	30	31	32	33	34	35	36	37	38	39	40
907	1219	1625	2158	2867	3823	5126	6913	9367	12728	17308	23513	31876
27	33	37	44	51	59	70	79	95	106	128	143	172
440	593	794	1057	1408	1882	2528	3417	4636	6311	8590	11685	15852
467	628	831	1101	1459	1941	2598	3496	4731	6417	8718	11828	16024

	···10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	···1	2	4	7	11	16	22	29	37	46	56	68	84	107	141	191	263	364	502
$k=9$	···1	2	2	3	3	4	4	5	5	6	6	8	8	11	11	15	15	20	20
	···0	0	1	2	4	6	9	12	15	20	25	30	38	48	65	88	124	172	241
	···1	2	3	5	7	10	13	17	21	26	31	38	46	59	76	103	139	192	261

29	30	31	32	33	34	35	36	37	38	39	40	41
686	926	1234	1626	2125	2765	3596	4690	6148	8108	10754	14326	19132
26	26	34	31	45	45	60	60	80	80	106	106	140
330	450	600	796	1040	1360	1768	2315	3034	4014	5324	7110	9496
356	476	634	830	1085	1405	1828	2375	3114	4094	5430	7216	9636

表 2

$k=2,3,4,5,6,7,8,9$ 时 $r_n^{(k)},s_n^{(k)},u_n^{(k)},v_n^{(k)}$ 的值.

式 3 则又使我们能把 b 与 c 合并，经过这些处理之后，状态是

$$* \ * \ * + * \ * .$$

即便不知尼姆值，你也可看出，(唯一的)好办法是击倒 d 列或 a 列.

图 5 是一个二柱滚球游戏专用的辅助轮盘，它可以帮助我们方便地读出八列或少于八列二柱滚球游戏状态的尼姆值，条件是我们已知开勒司游戏或道森氏开勒司游戏(对于道森氏象棋，只要把尼姆值左移一位就行了)中排成一行的 n 个木柱的尼姆值.

n	0	1	2	3	4	5	6	7	8	9	10	11	12
开勒司	0	1	2	3	1	4	3	2	1	4	2	6	4
道森氏开勒司	0	0[1]	1	1	2	0	3	1	1	0	3	3	2

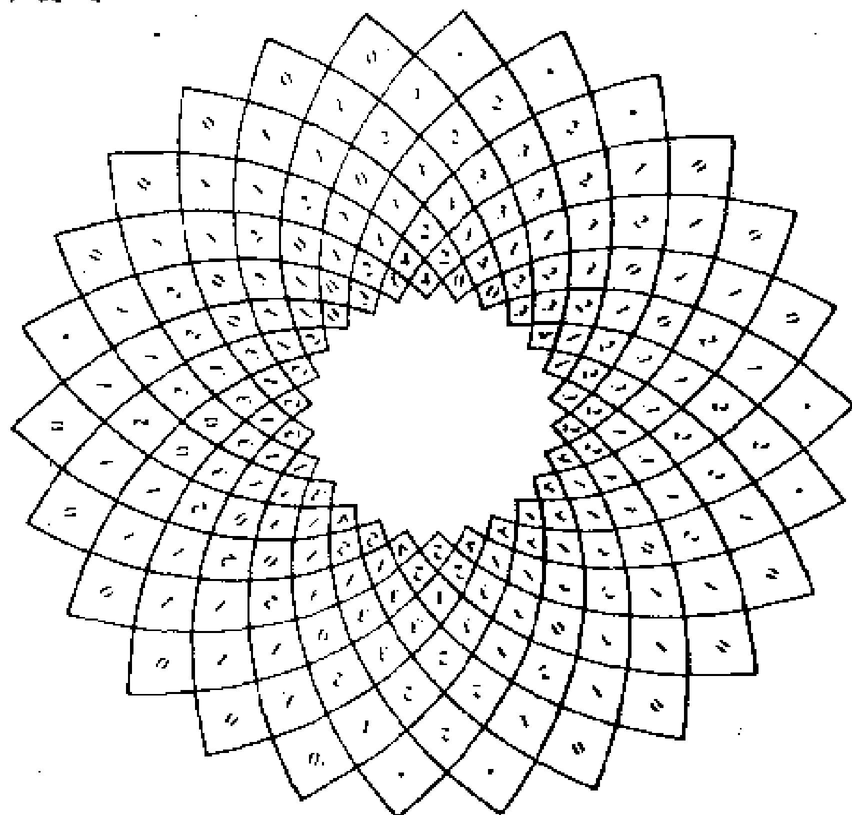


图 5

推算小型二柱滚球游戏状态的尼姆值的轮盘.

[1] 原注：请注意，在道森氏开勒司游戏中，单个木柱必须继续保留着.

譬如说，你想求出状态 * * * 0 0 0 * * 的尼姆值.

我们首先在最外缘的一圈中寻找这个配置状态(从12点钟到3点钟的位置)，然后按首、尾两个星号所在的格子循螺旋方向读入，在两者相交的格子中写着数字4，它就是尼姆值.

在图2中，道森氏象棋游戏的最优策略又是什么呢？我们的劝告是，请让你的对手先走，这是一个 P 位置(后发制人的位置[1])，其尼姆值为0.

参考文献

1 Austin, Richard, and Guy, Richard. 1978. Binary sequences without isolated ones. *Fibonacci Quart.* 16.

2 Ball, W. W. Rouse, and Coxeter, H. S. M. 1974. *Mathematical Recreations and Essays.* 12th ed. Toronto: Univ. of Toronto Press. pp. 36-39.

3 Berlekamp, E. R.; Conway, J. H.; and Guy, R. K. 1980. *Winning Ways.* New York: Harcourt, Brace Jovanovich.

4 Bouton, Charles L. 1901-2. Nim, a game with a complete mathematical theory. *Ann. Math.* Princeton(2)3: 35—39.

5 Conway, J. H. 1976. *On Numbers and Games.* New York: Academic Press.

6 Dawson, T. R. 1934. Problem 1603. *Fairy Chess Review.* p. 94.

7 ______. 1935. Caissa's Wild Roses. *Fairy Chess Review.* p. 13.

8 Dudeney, H. E. 1958. *Canterbury Puzzles.* N. Y.: Dover. pp. 118-119, 220.

9 Gardner, Martin. 1960. *More Mathematical Puzzles of Sam Loyd.* N. Y.: Dover. pp. 5, 122.

10 ______. 1974. Mathematical Games: Cram, crosscram and quadraphage: new games having elusive winning strategies, *Sci. Amer.* 230 2: 106.

11 Green, Thomas M. 1968. Recurrent sequences and Pascal's triangle. *Math. Mag.* 41: 13—21.

12 Grundy, P. M. 1964. Mathematics and games. *Eureka.* 27: 9-11.

---

[1] 译者注：直译是：在此状态以前的一方可赢，其行文非常晦涩，不易为我国读者理解，故改译如上.

13 Grundy, P. M., and Smith, C. A. B. 1956. Disjunctive games with the last player losing. *Proc. Cambridge Philos. Soc.* 52 : 527—533; M. R. 18 : 546.

14 Guy, Richard K., and Smith, Cedric A. B. 1956. The G-values for various games. *Proc. Cambridge Philos. Soc.* 52 : 514—526; M. R. 18 : 546.

15 Hardy, G. H., and Wright, E. M. 1960. *An Introduction to the Theory of Numbers*. 4th ed. Oxford : Oxford Univ. Press. pp. 117-120.

16 Jarden, Dov. 1966. Recurring Sequences. *Riveon Lematematika* 2nd ed. 86-91.

17 ______. 1946—47. Third order recurring sequences. *Riveon Lematematika* 1 : 74; 1952—53 6 : 41—42.

18 ______ and Katz, A. 1947-48. Table of binary linear third order recurring sequences. *Riveon Lematematika* 2 : pp. 54-55.

19 Loyd, Sam. 1914. *Cyclopedia of Tricks and Puzzles*. New York: Dover. p. 232.

20 Sloane, N. J. A. 1973. *A Handbook of Integer Sequences*. New York: Academic Press.

21 Sprague, R. P. 1935-36. Über mathematische Kampfspiele. *Tôhoku Math. J.* 41 : 438—444; Zbl. 13 : 290.

为寂寞无聊的数学家提供消遣的一种单人纸牌游戏

● 埃因侯温[1]工程技术大学

□ 德·布鲁因(N. G. de Bruijn)

一般来说，为了自娱目的而去玩单人纸牌游戏实在有点可怜巴巴. 从桌子上一堆乱牌出发，按某种预先规定的模式将其重新排列后，我们将试图找到一系列正规走法，使纸牌的状态达到某种预定目标. 玩牌的步法通常是不可逆转的，因此，常常会到达一种僵死状态，这时，任何一种移动都不可能. 如果我们有足够的好运道来达到目标，我们会感到某种程度的满足. 但若我们走进一个死胡同，我们将感到灰心丧气. 悲哀的原因倒不是由于命运有意同我们作对，而是我们在怀疑，失败是否不可避免. 我们感到伤心的理由是：也许是我们自己没有本事，把游戏搞糟了. 经常发生的情况是，我们已经记不起原来的状态，现在也已不可能回头追踪我们所走过的步数. 如果一个单人纸牌游戏具备完全信息（譬如说，所有的牌都是面朝上的），解决不了这类问题就会感到更加难堪，面上无光.

1. 庆幸自己的运气特别好，能够或多或少地通过偶然碰巧而解决难题，于是感到沾沾自喜，这样的想法绝不是数学家的习性. 如果问题有解，我们就希望通过什么手段把它寻出来，如果问题确实无解，我们就希望能够证明它不存在一种解答.

[1] 译者注：埃因侯温(Eindhoven)，位于荷兰南部.

数学家的目标是：他能够对某种给定的扑克牌状态进行理论上的分析，找出其解法或证明它无解，而不必去实际触摸一张纸牌. 很明显，只能对不太困难的单人纸牌游戏做到上述要求. 我们将建议用于这种研究的纸牌游戏有着两个整数参数，其数值大小取决于游戏者的胃口与能力，要求难易适中，恰到好处. 我们将把这种游戏命名为 $k\times n$ 椒盐卷饼. 这种游戏由来已久，特别对 $k=4$，$n=13$（详细玩法可参看本文第 5 节）更是源远流长，要追溯到它的起源是很不容易的. 本游戏的名称取得很新奇，之所以如此取名是由于椒盐卷饼是一种打结形状的饼干[●].

2. 为了做 $k\times n$ 卷饼游戏，我们需要事先准备一套具有 k 种花色，每种花色有 n 张牌的扑克牌. 为了把事情搞得简单一些，可令 $k=4$，并把四种花式分别记为 S，H，D，C（即黑桃、红心、方块与梅花），每套花色里头牌的点数是从 1 到 n（1 点通常称为“爱司”），这么一来，我们的记法便是 S1，…，Sn，…，C1，…，Cn.

在洗好牌以后，我们把扑克牌面朝上地放在桌上，排成一个 4 行、n 列的矩形. 接着，我们取走所有的爱司，并按照 S1，H1，D1，C1 的顺序放在长方形的各列之前，使它成为新的一列. 这样，我们就得到了一个有着 4 个空白地位的 $k\times(n+1)$ 矩形，下文将称之为空位，

♠1	—	♥4	♥2	—
♥1	—	♥3	♣2	♣4
♦1	—	♦4	♠3	♠2
♣1	♦2	♦3	♣3	♠4

图 1

每个空位所留出的就是一张扑克牌的大小. 要注意的是，即使 4 个空

● 译者注：类似于我国的“巧果”.

位都在最后一列，我们仍然称之为空位. 让我们设想，在这些扑克牌的外缘围绕着一个 $k\times(n+1)$ 矩形框架，藉以区别空位与浩瀚无垠的外部世界. 在游戏的进行过程中我们绝不容许把扑克牌移到框架外面. 作为 4×4 卷饼的一种状态，我们给出图 1.

现在我们开始做游戏，所谓走一步的意思是指把一张牌放到空位中去，但只有当和它同一花色且比它点数小一号的那张牌正好位于空位的左边贴邻时才能这样做. 因此，对图 1 的状态来说，牌 S2，H3，H2，D2 中的任一张牌都可以用来作为第一步的移动对象. 走了第一步之后，我们将会填满一个空位，同时又产生了一个新的空位. 这一新的空位又将要求我们走第二步. ……这样反复进行，我们希望能达到最后的状态：

♠1	♠2	♠3	♠4	—
♥1	♥2	♥3	♥4	—
♦1	♦2	♦3	♦4	—
♣1	♣2	♣3	♣4	—

图 2

我们把此种状态称为*目标*.

如果存在着一系列的步法得以把某一状态引向目标，则这个初始状态称为是*可解的*(目标本身也称作可解的). 如果不存在这种序列，我们就称之为不可解. 如果一个状态有别于目标，而且连一步也动弹不得时，就称之为*僵死的*. 当然，僵死状态肯定不可解.

这里，我们已对 $k=4,n=4$ 的情形讲解了游戏规则，但对所有其他情况，游戏规则也都是相同的.

3. 上文所提到的图 1 状态是可解的. 例如，按照下列走法(每一步都用搬动的扑克牌来表示)：

S2,S4,C4,C3,D4,D2,D3,C2,H4,H2,S3,S4,D4,C3,C4.

现在让我们改变记法. 显然,图 1 中的各列在我们的游戏中不起任何作用,所以我们只要单独一行再辅以分隔记号便可以把状态完全记录下来,例如图 1 可记成:

S1—— H4H2—— * H1—— H3C2C4 * D1—— D4S3S2 * C1D2D3C3S4 *

4. 4×2 与 4×3 卷饼是非常乏味的游戏,然而 4×4 卷饼却有可能搞得很为有趣. 4×5,4×6 通常还可一玩(按照第 2 节所阐明的意义来看问题).

对很大的 n 值来说,笔者还缺乏足够的经验,因此不敢断言一位中等水平的游戏者在既不触摸实际纸牌,又不利用纸笔或计算机的情况下,能否解决问题中的大多数情形.

5. 设 $p(k,n)$ 是一个随机的 $k\times n$ 卷饼状态有解的概率. 看来似乎难以正确地定出 $p(k,n)$ 之值,除非 $n\leqslant 2$,如 $n=2$,则唯一无解的状态是那些轮转封锁态. 所谓轮转封锁态就是各种花色中某一子集的轮转排列,例如 H→D→S→H,而(D1H2),(S1D2)与(H1S2)是各对相邻之牌. 如 $k=4,n=2$,则具有一个 3 数轮转的情况为 40 种,一个 2 数轮转的情况为 168 种,一个以上 2 数轮转的有 6 种,总而言之,无解状态共有 214 种,而状态的总数为 1680 种.

可以作出估计,$p(4,4)$的值大约是 0.45 左右,笔者陆续地玩过好几百个 4×4 卷饼游戏. 在所有场合他都能够判定一个状态是否有解(虽则他时常偏离所谓"不触摸牌"的规则而使用了一种"安全步法"——详见下文第 10 节). 另外,已有一个计算机搜索程序处理了 2473 个状态,其中可解的有 1123 个,不可解的有 1350 个.

笔者对 $n=5$ 与 $n=6$ 的经验很不够. 一个极为粗糙的估计是:$p(4,5)$的值大概属于 0.1 这一数量级.

4×13 卷饼游戏看来仍然有着相当数量的可解情形. 通常的玩法是,不一定指望能达到目标而是尽可能地接近目标. 游戏者希望在第一行中获得一个很长的序列 S1,S2,…,Sp. 类似地在第二、三、四

行中获得 H1,H2,…,Hq,D1,…,Dr,C1,…,Cs. 当局势变成僵死状态时,他把不在这些序列中的牌统统拿掉,然后重新洗牌,再次填满矩形,并在 Sp,Hq,Dr,Cs 的右方紧邻分别留出一个空位,然后重新开始做游戏. 在第二轮之后,也可能还会有第三轮,如果游戏者能在第三轮达到目标状态,他就被看作是成功的.

但不时出现第一轮就达到目标状态的情况. 让我们告诉大家,发生这种事情的机会大约是全部情形的百分之一. 但这不过是对 $p(4,13)$的一个极低估计,因为无人知晓究竟有多少可解的情况被搞糟了. 通过此种经验数据,人们可以猜想,当 n 趋于无穷大时,$p(4,n)$也许不是按指数规律迅速衰减的.

6. 在第 5 节中我们提到,4×4 卷饼游戏曾在计算机上作过模拟运行. 程序是在研究卷饼游戏的"图"的基础上建立的. 图上的点是一切可能状态. 如能走一步动作,从状态 P 到达状态 Q,则在图上就可以用一条从 P 到 Q 的有向边来表示. 若 P 为一个状态,则 S(P)就表示从 P 出发经过有限步数所能到达的一切点的集合. 我们的问题是:目标状态 P_0 是否属于 S(P).

对任意一个给定的P,在计算机的存储器中易于建立集合 S(P),这只要考虑,从原有的点出发,经过一步动作之后将能得到哪些新点. 如果 S(P)中不再可能找到新点(这意味着状态 P 是不可解的),或者已达到了目标状态 P_0,这时程序就将停止执行. 在后一种情况,我们就说 P 是可解的,因此我们并不需要知道完整的 S(P).

如果状态 P 是不可解的,则集合 S(P)中元素个数很少超过 60;话虽如此,但也有反例,曾发现过的最大个数是 380. 对可解的 P 来说,在达到 P_0 之前所经过的点的个数一般在 150 左右,但也有大到 802 的. 不过 S(P)的大小并不是饶有兴味的,如果它太大或太小,那么通常就可以用心算来计算 P.

在卷饼游戏中,从一个状态到另一个状态可能有许许多多不同的走法,因为,有可能前后各步互换走法而彼此不受影响. 这意味着,为了要寻找一系列步法,而去设计一种逆推法计算机程序,也许是不

甚合适的.

7. 在 4×4 卷饼游戏中,从任意可解的初始状态 P 出发,纯粹凭藉思考办法来找出解答一般说还不算太难,虽则有时候,人们也会遇到一个刁钻古怪的情况而碰壁. 但若初始状态是不可解的,我们又将如何来说明? 想把第六节所说的 $S(P)$ 完全地决定出来,对一个不是计算机的脑袋来说简直是不可能的,但是,有许多其他办法可以证明不可解性.

其中一种办法是轮转封锁状态. 第五节中已讲过它的最简形式. 但通常却要复杂得多,有如下例. 我们注视着某张牌 c_1,然后我们看到,在第一次移动这张牌之前,必须先移动牌 c_2. 接着,我们又看到,在移动 c_2 之前,不能移动牌 c_3. 因此,在任何一种解法中,必然有一时刻要首先移动牌 c_2,而 c_1 与 c_3 则尚未移动. 继而,我们又发现,在移动 c_3 之前必须先得挪动 c_4,如此等等. 在这种推理过程中,我们将会看到,在我们心里头正在考虑的那个时刻,尚未移动的牌越来越多,而且还在一个劲地继续增长. 如果或迟或早,我们跌进了一个死循环,发现在移动所有各张牌之前,必须先得挪动 c_1,则我们就证明了原先的初始状态是不可解的.

玩这种游戏的数学家无疑将会发现一只口袋,里面装着许多增强其证明能力的各种新手段.

这些新手段有的只能意会,难以言传. 它们常常是前进推理与后退推理的混合物. 后退推理在刚才讨论轮转封锁状态时已经说明过. 至于前进推理则是要看到从初始状态迈出的各种不同的第一步将会产生的一切后果;譬如说,从 P 走一步可以到达 P_1,P_2,P_3 或 P_4 中的一个,而若 P_1,P_2,P_3,P_4 全部都是不可解的,则 P 也不可解. 我们必须慎之又慎,尽量避免大规模地混用前进与后退推理,因为那将很容易得到一个如愿以偿,但却是实质不可能的证明,而这会比根本没有证明坏得多. 通常,如果我们代之以第 8 节将要讲到的截短定理的话,这种伪证可以避免.

在求解方面的挫折将有助于得出一个无解的证明,而在寻求不

可能性证明方面的挫折将有助于找出一个解. 总而言之,这一切正体现了数学家工作的灵活性.

8. 设 p 是一个整数, $1 \leqslant p \leqslant n$, 则 $E_S(p)$ 是点数不大于 p 的黑桃扑克牌的集合(即 $\{S1, S2, \cdots, S_p\}$). 类似地我们可定义 E_H, E_D, E_C, 并令

$$F(p,q,r,s) = E_S(p) \cup E_H(q) \cup E_D(r) \cup E_C(s).$$

这样的集合 $F(p,q,r,s)$ 称为主干; 若 $k \neq 4$, 其定义也与此相仿.

若 P 是 $k \times n$ 卷饼游戏中的一个状态, F 是一个主干, 则在 P 中把不属于 F 的所有扑克牌统统拿掉后所余之状态称为"P 以 F 截短". 从第 2 节所说的意义上来看, 它不再是一种状态(除非 $F = F(n, n, n, n)$), 然而我们仍旧可以在被截短后的状态下做游戏, 并保持第 2 节所说的步法定义不变. 我们约定, 不论 F 是何等样子的集合, 截短后的 $k \times n$ 卷饼游戏仍然要在原先的 $(k+1) \times n$ 框架(因而 $k \times n$ 卷饼游戏与 $k \times (n+1)$ 卷饼经 $F(n,n,n,n)$ 截短后的游戏不是同一个东西)内去玩. 我们把 P_0 被 F 截短后的状态视为被截短的卷饼游戏的目标, 这里 P_0 是原先的卷饼游戏之目标. 同样地, 一个被截短的状态称为是可解的, 如果有一系列的走法, 能把它引向新的目标状态的话.

设 P_1、P_2 均是状态(或截短状态), 而且走一步就可从 P_1 变为 P_2. 设 F 是一个主干, 而 P_i^* 为 P_i 经 F 截短后的状态. 如果移动的牌属于 F 这个集合, 则同样的一步可用来使 P_1^* 变为 P_2^*. 如该牌不属于 F, 则我们有 $P_1^* = P_2^*$. 将此引用到一系列的变换步法, 若 F 为固定时, 则可得到下面的定理, 它可视为检验可解性的一项必要判据.

定理 *若 P 是可解的, 则 P 的任一截短都是可解的.*

9. 第 8 节所说之截短定理的最简单应用是某一状态在截短后变为僵死的情况. 例如,

S1D4S3D2H2 * H1D3—H4— * D1H3C2S2— * C1C4S4C3— *

是不可解的，因当它以 F(1,3,3,1)截短后，将变为

S1———D2H2 * H1D3 ——— * D1H3——— * C1——— *

到此地步，它已经一步动弹不得.

在较复杂的场合，我们证明截短后的状态是无解的，但并不是僵死状态. 或者我们证明从我们的初始状态出发，走每一步都将到达一个状态，对于这个状态，我们能选一个不可解的截短，如此等等.

10. 除非从一个可解状态走成一个不可解状态，否则都称作是安全的步法. 对一个不可解状态来说，一切走法都是安全的，但对许多可解状态来说，我们却有着不安全的走法.

通常我们可以相当有把握地证明一步动作是安全的. 设在某一状态 P 中，牌 c 的位置是 p. 如果能够同时满足下面的 1、2 两个条件，则把 c 从 p 搬到 q 的动作一定是安全的：

1. p 的右方紧邻的那个位置要末落在框架之外，要末它已被同一花色中点数高一号的牌所占据，要末在同一花色中已没有点数更大的牌.

2. 将 P 转变为目标 P_0 的任何一系列步法，都不可能在 q 的位置上放下一张不是 c 的牌.

条件 1 的目的是要保证对位置 q 上的牌 c 没有别的用途；条件 2 则保证牌 c 在位置 q 上不能造成任何损害.

至于条件 2，我们想说：某些牌是永远到不了某些位置的. 例如，如果我们有一行牌 S1H3C2D4S3D2D3，则我们可以断定，在考察了左边的位置以后，目前由 D2 占据的位置，今后可能的候补者只能是 S6，H7，C5，D6 与 S4.

11. 安全步法的另一种来源如下. 如果在一系列的走法中，移动的牌都是方块，而且在走法的最后一步，所有的方块都已处于适当位置(即它们在最后的目标状态所应到达之位置)，则这系列走法由安全步法组成. 在上面这句句子中，如果在出现“方块”这个字眼的地方代之以“方块与黑桃”或者“方块、黑桃与梅花”，论断依然正确无误.

以上这些论证均易于从截短定理推出.

12．如果某一状态 P 只存在唯一可能的一步走法，则这步走法显然是安全的．

现在我们设想有一个存在两种可能走法的状态 P，有两张牌 c_1，c_2 可打，并假定 c_2 不是与牌 c_1 花色相同，且其点数比 c_1 多一点或少一点的牌（这个条件保证了两步走法互不干扰，也就是说，在移动了牌 c_1 之后，原先想挪动牌 c_2 的走法仍然是可能的，反之亦然）．进一步我们再假定，c_1 将把 P 转变为一种状态，其时 c_2 将是唯一可以利用的一步．于是我们就可以说，c_2 是关于状态 P 的一个安全步法．

证明安全性的这一简单想法可以推而广之，既包括互不干扰的两步，也包括互不干扰的步法序列．寂寞的数学家无疑会发现许多类似论据．

最后我们想提一下，在任何状态下，要证明一个步法是安全的，只要能够证明所有其他走法都是坏的（这就是说，将导致不可解状态）就行．

13．在第一节中我们曾说过，无需触摸一张牌就可分析本游戏．但有时难免遭受挫折，因此修改一下规则并非不合情理，即允许做游戏者能走出一些步法，只要他能够证明这些步法都是安全的．

做游戏者能加以采用的一种记忆方法与第 7 节中提到过的那种封锁状态分析有联系．分析某一张牌第一次搬动会发生些什么情况时，游戏者可以在他认为直到那时无法搬动的每一张扑克牌上面放一个小分币以作为记认．

14．迄今为止我们的卷饼游戏满足如下条件：

1．每一套花色具有同样张数的牌．

2．在框架中（见图 1），所有的行都具有同样长度．

3．空位的个数等于花色的种数．

我们有时也能忽视以上三个条件而照样有游戏可做，使在前面章节里所讲到的一切东西都依然保持有效．作为一个实例，我们提出

S1D4D3——S3S5 * H1H2S4D2 * D1——S2H3 *

其解法是 D2，S5，S4，H3，D4，S2，D3，S3，D4，S4，S5．

15. 我们提供一些练习，选自各种随机状态的汇编. 打上"☆"记号的习题是相当困难的：笔者的技巧口袋里似乎还缺少锦囊妙计. 下面的 1 至 9 题是 4×4 卷饼游戏，10 至 16 题是 4×5 卷饼游戏. 解答就附在下面.

1. S1D2 —— D3S3 * H1 —— C2H2D4 * D1 —— S4C3H4 * C1S2H3 —— C4 *

2. S1 —— H2H3S2 * H1H4S4 —— D2 * D1 —— C3S3C2 * C1D4D3C4 —— *

3. S1C3 —— C4H4 * H1 —— D4H2C2 * D1D3S4S2H3 * C1 —— D2 —— S3 *

4. S1S2D4C3D3 * H1C4 —— H3H4 * D1S4C2H2 —— * C1 —— D2 —— S3 *

5. S1C4C2D3 —— * H1H2H3H4 —— * D1S4S3C3D4 * C1D2 —— S2 —— *

6. S1C4H4 —— H2 * H1D3D4C3S4 * D1S3 —— H3 —— * C1S2 —— C2D2 *

7☆. S1D3C2 —— H2 * H1S3D4D2 —— * D1H4C3H3C4 * C1 —— S4S2 —— *

8. S1C2C3S2 —— * H1 —— D2H4 * D1D3H3S4H2 * C1C4S3 —— D4 *

9. S1D4S4S2H4 * H1C4D3H2C2 * D1 —— C3 —— H3 * C1S3 —— D2 —— *

10. S1D4C4D3 —— C5 * H1 —— D5 —— H5H2 * D1C2S3H4H3S4 * C1S2S5 —— C3D2 *

11. S1S5D2H2 —— D4 * H1H4S4S3 —— D5 * D1H5C5C2D3 —— * C1S2 —— C3H3C4 *

12. S1C4D5D3S5C2 * H1 —— D4 —— C3S2 * D1C5H5H3 —— H4 * C1D2H2S4 —— S3 *

13. S1 —— C4H4C5D4 * H1D3C2S2H2H5 * D1D5 —— D2 ——

H3 * C1C3S3S5　　S4 *

14*. S1S5S2D5H3D4 * H1——C2D3S3H4 * D1——H2C4H5C3 * C1C5——S4D2——*

15. S1C4——D2C5S4 * H1S5D5H5D4D3 * D1H2S2H4——C3 * C1S3H3——C2——*

16. S1C4H2D4——D2 * H1C3C5D3S5H5 * D1——H4S4——C2 * C1　　D5H3S3S2 *

解　　答

1. 可解：D3,D4,H3,S3,S4,D2,D3,H2,S2,S3,C2,C3,D4,S4,C4,H3,H4.

2. 不可解.以 $F(1,\cdot,3,3)$ 截短之,走两步后即僵死.

3. 不可解.以 $F(4,1,3,4)$ 截短之,于是 D3 是一步安全走法,接下去,D2,C2,C3,S2,S3 是一系列安全走法,最后导致一个僵死状态.

4. 不可解.以 $F(4,1,1,1)$ 截短之,走一步即僵死.

5. 可解：D4,C4,D3,C3,S4,S3,D2,D3,C2,C3,S2,S3,S4,C4,D4.

6. 不可解.H3 与 D4 相互封锁.

7. 不可解.

8. 不可解.以 $F(2,1,1,4)$ 截短之.

9. 可解：C4,H2,D4,S2,S4,S3,S4,D2,C2,C3,C4,D3,D4,H3,H4.

10. 不可解.搬动 H2 是安全的,其后 14 步都是唯一的,最后导致一个僵死的盘面.

11. 可解：D4,H3,H4,H2,C4,C5,D3,C3,S3,S4,H3,H4,H5,D4,C4,D2,D3,S5,S2,S3,S4,C2,D4,S5,C3,C4,C5,D5.

12. 不可解.以 $F(4,1,1,4)$ 截短它之后,走两步即导致一个僵死的盘面.

13. 不可解.搬动 S2 是安全的步法,因为搬动 D3 将导致一个不可解状态(用 $F(1,1,5,1)$截短之)).根据同样的理由,C3,C2,D4 也是安全的,其后的 13 步都是唯一的(除掉第 10 步与第 11 步可互换之外),最后导致盘面僵死.

14. 不可解.

15. 可解:H4,C3,S3,S4,S5,C5,C2,H2,D3,D5,H3,C3,H5,H4,D4,H5,C4,C5,S2,S3,D2,D3,D4,S4,S5,D5.

16. 可解:D2,C2,S5,D4,H3,H4,H5,D3,D5,C3,C4,H2,S2,S3,C5,H3,H4,S4,S5.D4,D5,H5.

一个Hex问题的若干评注

● 巴黎大学

□ 克劳特·贝尔热(Claude Berge)

读过加德纳文章(请参看《科学美国人》杂志1957年7月,1957年10月,1975年7月与1975年12月各期)的读者也许能理解Hex游戏的深刻与美妙.这个游戏是由丹麦的皮特·海因(Piet Hein)与美国的约翰·纳希(John Nash)差不多同时推出,它对数学家们特别富有吸引力.两位局中人轮流下子,每次须在图1所示的菱形棋盘的空穴中放置一枚尖头棋子.先手用的是黑色棋子,如能从东到西构成一条黑棋长链,他就算是赢家.他的对手则企图用白棋构筑一条从南到北的长链.棋盘的最优形状似为14×14,由于反证法表明在先手方面存在着一种获胜策略,因此他的对手有权要求他的第一着棋子只能下在棋盘的某一限定区域内.如果两位局中人不是功力悉敌的,那么形形色色的不利条件可加诸棋力强的一方以使得获胜机会均等,尽管如此,该游戏的美学情趣却基本未变.

我想提出一个Hex问题以作为对马丁·加德纳诞辰的献礼.一个问题的引人之处也许是:第一着是独一无二的,或者第一着是荒谬绝伦的,也可能着法的构思是出人意外的,如此等等(我们不妨回忆一下山姆·洛伊德,这位一切时代的象棋排局大师,他曾宣称其主要目标是创作一种非常独特的象棋问题,其第一步走法与千分之九百九十九人的想法恰恰相反).

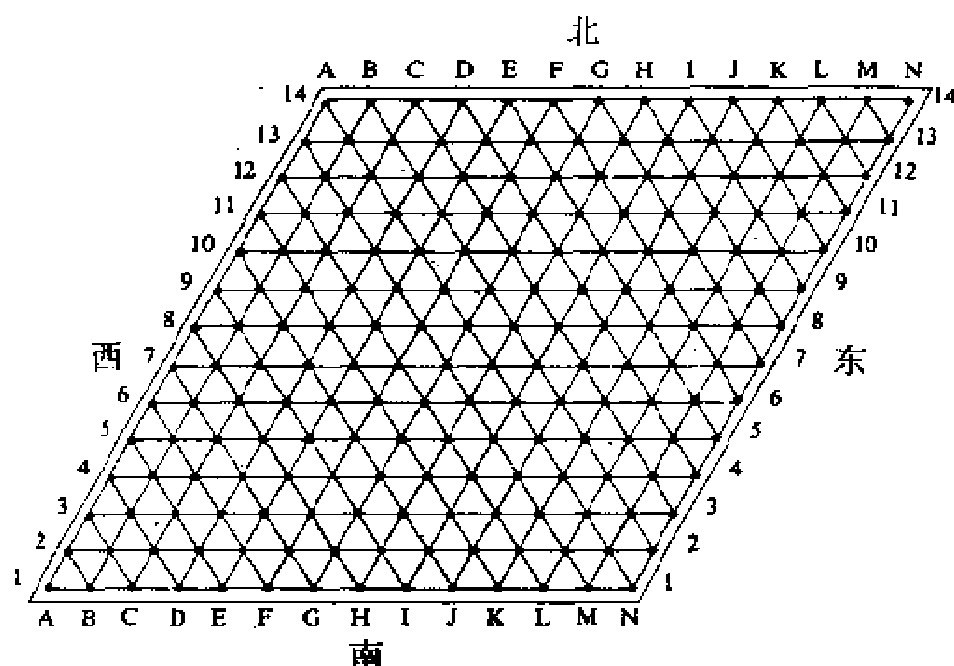


图 1

空白菱形棋盘.

这里,为了避免不必要的纠缠,以及允许一些中介步法,我们不打算采用通常的判别标准——即要求关键性的一步必须是独一无二的.我们也无需要求棋盘上有同样子数的白棋与黑棋,因为前已说过,对棋力强者可以事先设置若干不利条件,一个棋局也可从让先开始…….我们要提出的问题其难度并不在于盘面复杂或者变化多端,因为白方可以在J9或M8下子,从而可以相当轻易地完成一条自南至北的通路,由于黑方不可能同时在这两个空位下子,因此他看来马上就得输棋.我们的问题是黑方怎样走法才能取胜!

对于棋势所作的简单分析表明:通到西方一侧的黑子可以被分割为两块,但通到东方的黑子则不然.事实上,这两群将被两堵不相连接的白棋构成之围墙所完全包围.其中的第一个是:

A11,B10,B11—C10,C11,D11—D10,E10,E11—F10,
F11,G11—G10,H10,H11—I10,I11,J11—J10,K10,
K11—L10,L11,L12—M11,M12,N12,N13,N14.

第二个是：

A8，B8，C8，C9—D8，D9，E8—E9，F8，F9—G8，G9，H9—H8，I8，J9—J8，J9，K9—K8，L8，L7，L6—M6，M5，N4，N3，N2，N1.

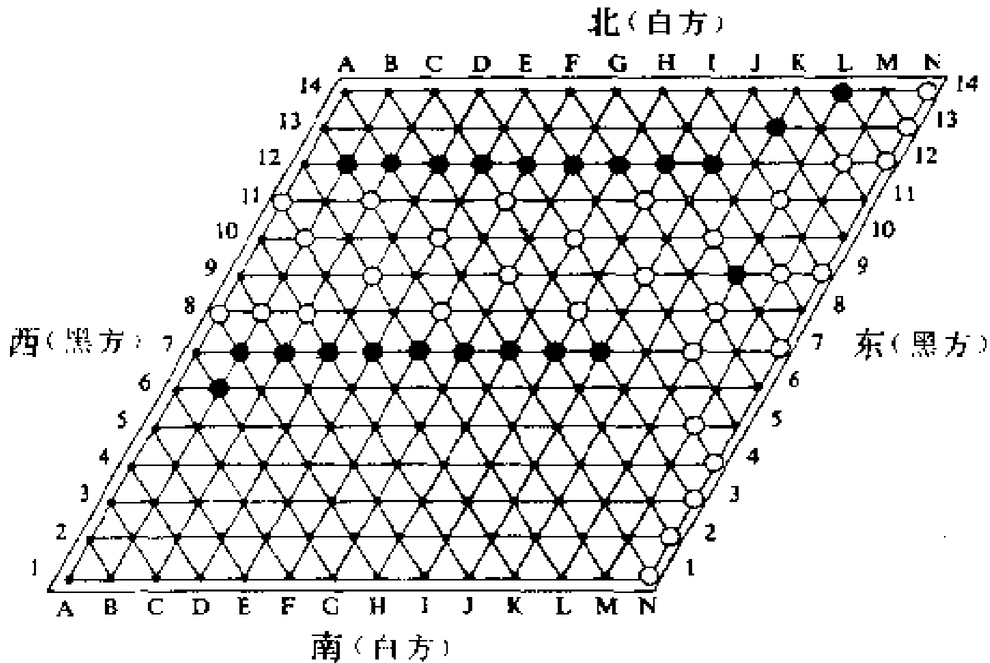


图 2

黑方怎样走法才能获胜？

你会发现，没有一个紧要的空白位置（例如C9，D8，…）同时属于第一堵墙和第二堵墙. 因此，不可能指望会有任何一种组合来打破“至少是一堵”白墙的封锁. 因而谁会猜得到唯一的一颗黑子L9竟能开辟出一条贯穿东西的新通路？在走了中间的几步黑子M6，M11与L10之后，黑子所成之长链将是：K9，J10，I10，I9，H9，G10，F10，F9，E9…．

利用某些有关集的组合性质（所考虑之集合乃是关键空位的集群）的具有相当深度的定理来解决一些Hex问题是很有意思的. 人们不可能忘掉这样的事实，山姆·洛伊德的一个包含对偶性的著名棋局（名叫“彗星”）被一位懂得二分图的König定理的一位数学家轻

而易举地解决了.此外,在象棋理论中,马赛尔·邓凯普(Marcel Duchamp)与阿尔柏斯塔特(Alberstadt)的共轭格子理论乃是图的代数同构理论的美妙应用(这两个图都是由王棋(king)的步法来定义的).

某一种数学工具的应用也许是出人意料的,从而可对一种游戏增添若干新的兴味.Hex 势将作为一种最富有娱乐价值的游戏而长期存在,不论对数学家还是对门外汉都将如此.

一种"瞎子打仗"残局棋戏

● 斯坦福大学

□ 吉姆·博伊斯(Jim Boyce)

"瞎子打仗"棋戏是国际象棋的最有趣变种之一：每一位局中人都想将死对方，他手中用的是普通棋子，也得遵守通常的棋规，可是却不知道对方的棋子究竟在哪里. 事实上，两位局中人都有一只秘不告人的棋盘，在其上记录着他们自己的兵将并猜测对方的棋子位置. 在这一棋戏中还有第三位参与者，名为棋戏的公证人，他手中有着第三套棋子，可藉以保持棋子所处的实际位置. 当轮到白方走棋时，他先向公证人提出一个可能走法，如果棋子的真实位置允许走这步棋，那么公证人点头认可，这一走法也就成了白方的正式一着. 如果情况不是这样，那么白方必须继续试探其他走法，并征询公证人意见，直至某一走法成为正式为止. 接下来就轮到黑方走棋，而棋戏就按照此种方式在进行. 如某方的任意正式一着使对方的王棋被"将"时，公证人必须向双方同时宣布. 此外还有其他下棋规则，例如兵卒走法以及吃子规则等(这些都与本文无关)，欲知进一步的细节，请参阅文献〔1〕.

本文要分析的残局是国王与堡垒对付国王的问题. 对普通国际象棋来说，这是极为初等的将死问题〔2〕，但在"瞎子打仗"棋戏的基本规则支配下，它决非轻而易举. 事实上，已有一些有经验的棋手在这个问题上虚掷了许多钟点而未获解决. 我们也建议读者在阅读下

文的解法之前不妨先试一试自己的手艺.

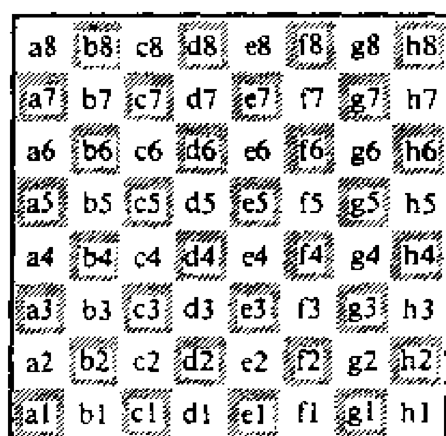


图 1

标准代数记法.

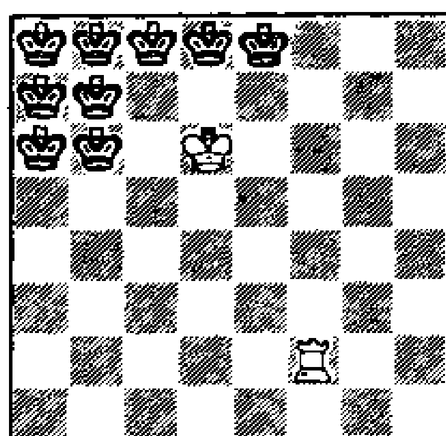


图 2

取作样本的对局状态.

走子及记法

在本文中，白方有一王棋与一堡垒，黑方只有一枚王棋. 走子记录由代数记法来表示，在棋盘上，黑、白格子相间，纵列的编号自左至右，从 a 到 h；横行的编号则自下而上从 1 到 8(见图 1). 所有的数字都是从白方的角度来看棋盘局势的，记下了他的棋子的已知位置，并对黑方王棋的每一种可能位置作了标志.

在我们讲过一、二个简单例子以后，上述规则与记法就会变得很清楚. 设想在图 2 中，白方想把他的王棋走到 d7 位置，如果公证人说这一步是不行的，白方的脑中就会出现图 3 那样的局势. 现在如果白方想把他的堡垒走到 e2，而公证人说了一声“将”，那就意味着黑方的王棋必定在 e8. 于是黑方从 e8 走了一步，下面再轮到白方走棋，此时棋盘上的形势将如图 4 所示. 白方将王棋走到 c6，在未落子前他已知道这步棋肯定是合法的. 在黑方走了一着后，白方就知道黑方的王棋不可能再在 d8 了. 如果它曾经在那里，此时必已走开. 然

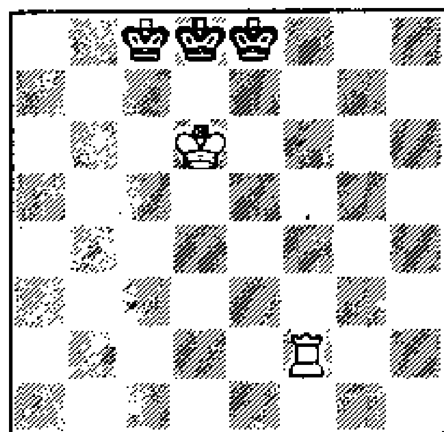


图 3
当公证人宣布白方王棋走到d7为不行时，盘面的形势.

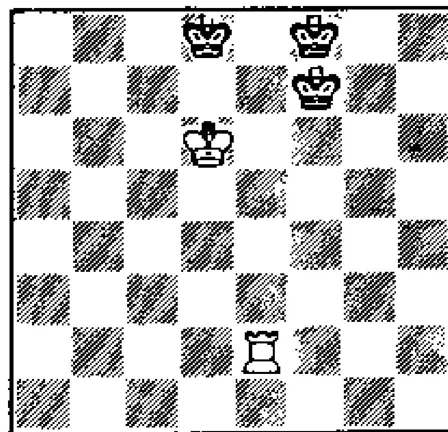


图 4
当堡垒走到e2"将"后的局势.

而，黑的王棋仍有可能在f8. 在f8的王棋也有可能必须走开，然而在f7的王棋可能走到f8.所以在黑方走了下一步后的棋局形势有如图5所示. 如果黑方王棋是在c8.白方即可把其堡垒走到e8，而把黑方将死. 图6表明了这一结果. 当然，如果白走了这一着，而黑的王棋是在f7或f8的话，则可把堡垒吃掉而造成和棋.

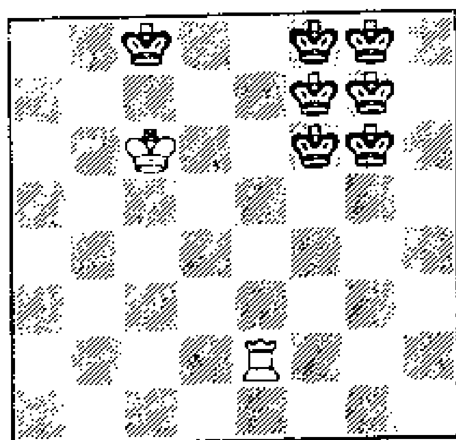


图 5
K—c6后的形势.

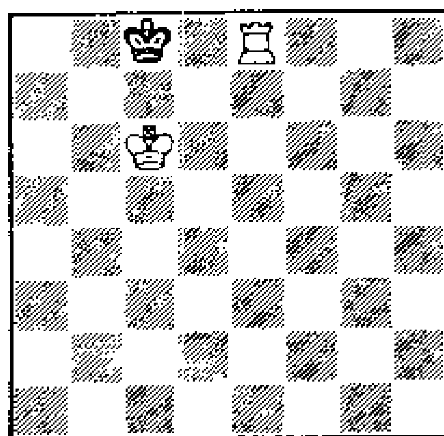


图 6
R—e8即可把黑棋将死.

自图 2 到图 6 的一系列走法可用代数记号记录如下：1. K—d7，R—e2＋，2. K—c6，3. R—e8#．现在请注意，第一点，R—e2＋之前没有数字，这是因为K—d7这一步是不能走的，白方仍需要寻找他的第一步走法．记号＋表示"将"，而记号#则表示将死．和棋（僵持状态或堡垒被吃掉）则用记号＝来表示，它有可能在出现＋或#的同一位置出现．

王棋、堡垒对王棋

我们将会看到，即使在"瞎子打仗"这种不利条件下，白方几乎还是能够赢棋的．黑方只能在某些初始条件下打成和局，其时，白方来不及运用他的王棋与堡垒，而黑方即已把堡垒吃掉．即使黑方有可能吃掉堡垒（设白方已知道这种形势而心怀鬼胎），一般来说，他也未必真能吃掉．因为白方如果真的把其堡垒置于黑方王棋的虎口之下，黑方还是需要猜测究竟要走到哪里才能吃掉堡垒．然而本文将只研究那些白方一定可以将死黑方的情况，即使黑方作了最高明的应付．

白方的计划可分为几个阶段．首先，他必须吃准，他的堡垒不会被吃掉．其次，他要设法走到这样一种状态，其时黑方王棋所能有的一切位置都落在一个矩形之内，该矩形的一个角是棋盘之一角，而其对角则是堡垒所处之位置．白方也需要把其王棋落在这个矩形之内以使黑方王棋远离其堡垒．然后他就可以驱赶黑方王棋，使它最后只能占据一条边上的几个方块．这样一来，要将死它就是相当容易的事情了．

通过考虑以下各种类型的状态，本文将说明一种简单（但非常缓慢）的办法来将死黑棋：

1．从堡垒的观点来看，双方王棋都处于棋盘的同一象限．

2．黑方王棋限于棋盘的一个或两个象限．

3．白方的堡垒是黑方吃不到的（或者可以使黑方吃不到）．

现在让我们把第1类型更详细地分析一下：

1a. 黑方王棋局限于一行(或一列).

1b. 黑方王棋局限于两行.

1c. 黑方王棋局限于三行.

1d. 黑方王棋局限于四行.

1e. 黑方王棋被限定的范围在四行以上.

1. 双方王棋都在同一象限. 在图 7 上,两只王棋都在堡垒的左上角. 白方企图将黑方王棋在棋盘上所占据之矩形越圈越小. 对此方针胸有成竹之后,白方就企图把他的王棋走到 e4,其堡垒走到 f3(他自然更乐于把他的堡垒推进到第四行). 棋戏将进行如次. 1. K—e4,R—f3(或者是:若 K—e4 是合法的一着,则可照 A 继续走棋,见下文),2. K—e4,K—e3,K—f5(或 B),3. K—e4,R—f4. 这一特殊的走法系列将可减少较小的一个象限的大小并得到一个类似于白方下了第一着之后的局势. 如果在第 2 步或第 3 步时走 K—e4 是合法的,则白方就可减少较大的象限的大小并获得一个与图 7 类似的局势. 对白方的试探,还可能有一些其他反应(通过公证人),如果 1. K—e4 是合法的,我们可能有 A: 2. R—f3 或 R—f3+, 3. R—f4,第二种情形将得出这样一种局势,有关象限的较小部分减少了,

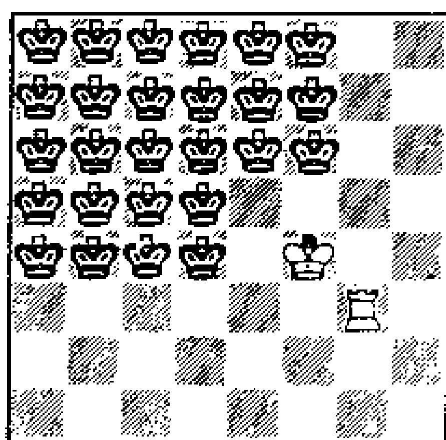


图 7

双方王棋在同一个象限

(由堡垒看)白方走棋.

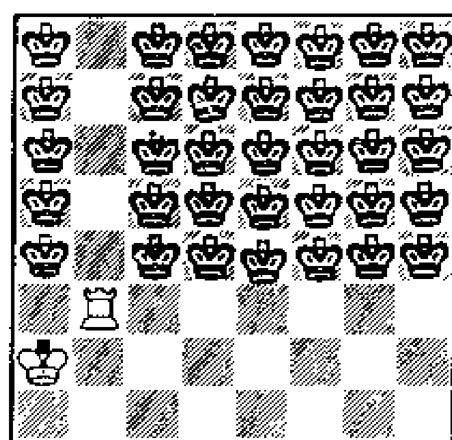


图 8

白方走棋.

然而黑方王棋可能位于两象限之一. 另外一种可能性是 2. …K—e3 是合法的. 这时 B 情况继续演变下去. 3. K—e4,R—f4,第一步是人们熟知的,第二步会缩小较小的象限的大小,但结果将产生这样一种局势：黑方王棋被局限于一个象限,而白方王棋却不在该象限. 从任何类似状态(除非较小的象限大小只是一)出发,也可采用与此类似的走法. 作为其结果,在任何情况下,可以证明白方的王棋与黑方的王棋将处于同一象限而不会增大较小一个的尺寸. 于是,白方即得以逐步限制对方王棋的活动范围,直至把它逼到棋盘的边上.

2. 黑方王棋在一个或两个相邻的象限内. 图 8 表明了一种局势,黑方王棋可位于图上任何两个象限之一. 白方企图达到一个类似与图 7 的状态. 例如,白方王棋在 b4,堡垒在 a3,黑方王棋在 4—8 行中的某一处. 第一步是移动堡垒(如有需要时,也包括王棋),使可能藏有黑方王棋的两个象限之一有一个长度为 2 的边. 由图 8,弈法可进行如下：1. K—b2, 2. R—c3. 然后白方将其王棋走入该象限, 3. K—b3, 4. K—b4,如黑方王棋进行干扰,白方只能撤回其堡垒. 于是双方王棋都将处于同一象限(有一边之长度是 2). 最后,白方把其堡垒走到边上, 5. R—a3. 在这一系列走法中,白方可以无需考虑会发生任何被"将"的情况. 如果黑方王棋限于一个象限之内,白方亦能应用同样的计划. 有时候,白方并不需要把其王棋与堡垒自边上越过第二与第三线. 在图 9 中,白方将其迂回到 c6,并将堡垒移到 c5,以获得一个第 1 节第 1 着后所产生之局势. 1. R—d4 也能产生一个使两只王棋都处于同一象限的局势,然而象限的较小者将会大些. 图 9 的棋戏可继续如下：1. K—d4, 2. K—c5, 3. K—c6, 4. R—b5.

3. 堡垒是安全的(或者能够安全). 如果王棋与堡垒在棋盘中部相连,黑方王棋不像上节那样受限制,那么白方希望用某种办法来限制黑方王棋. 有一种局势颇似第 1 节所说,堡垒位于角上,例如 h1,他一方的王棋在附近,例如 g2,于是白方把他的棋子向角上移动. 如果黑方王棋进行干扰,则白方就知道其位置是受了限制并可照第 1

或第 2 节进行. 如果不是这样，则白棋就走到角中，此时可遵循第 1 节的方式进行下去.

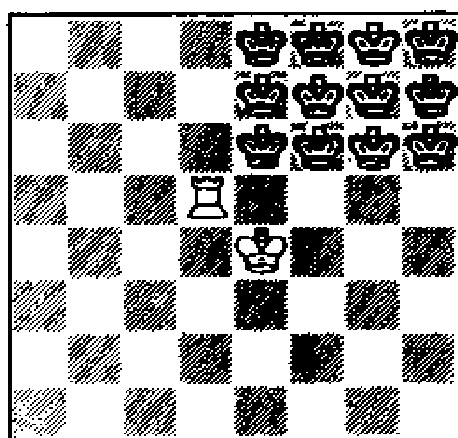


图 9

白方走棋.

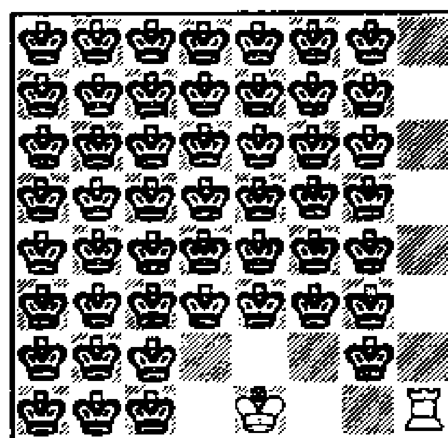


图 10

问题 1. 多少步救出堡垒?

如果堡垒有被吃掉的可能，则白方的第一件事情就是要保护其堡垒. 这通常很简单. 有时候，有好几种办法保护堡垒. 图 10 的局势里有一只未受保护的堡垒. 读者们能否发现所有的救它脱险之法? (答案见本文末尾.)

1a. 黑方王棋被限于一行.

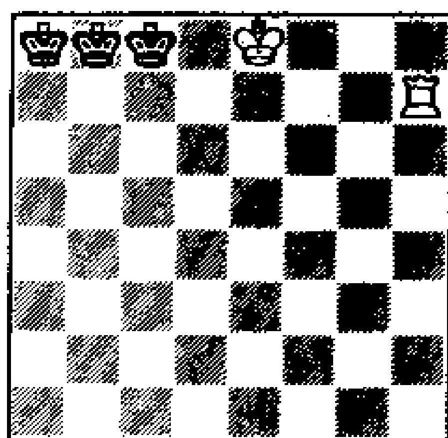


图 11

白方走棋.

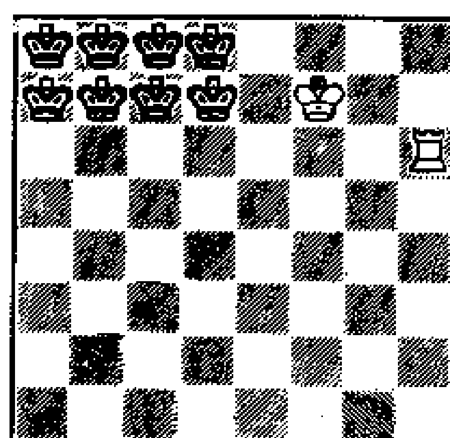


图 12

白方走棋.

上述分析不适用于两只王棋同处于一个象限，而这个象限其实是棋盘边上一行的情况. 图 11 表明此种态势. 白方很容易将死对方，只要把王棋逼回到角里，然后用堡垒来将它. 但白方必须小心，以免形成僵局. 此时棋戏可进行如下： 1. K—d8，R—g7，2. K—d8，3. K—c7（不是 3. K—c8＝），4. R—g6，5. R—a6＃. 当黑方王棋使白方王棋不能移动时，堡垒可以走棋，于是黑方王棋只好后退.

1b. 黑方王棋被限于两行. 如果黑方王棋已被限制于两行，则白方可改善 1 中所给出的方式. 在把黑方王棋赶进角上时，已经没有必要使堡垒随着白方王棋协同前进. 由图 12，弈法可如下进行：

1. K—e7，2. K—d7，K—d6，3. K—d7，4. K—c7，K—c6，K—d8，5. K—c7，6. K—b6，K—c8，7. R—a6＃.

如果 5. K—c7 不合法，白方可改走 5. …R—h7. 走两步后，7. R—h7＝将是一个严重错误，但很容易避免.（第 1 节里讲过的方式表明能削减黑方王棋的活动区域，没有提到的一点则是，这也可以与黑方僵持. 如同最后的例子那样，那是很容易决定的，白方将有一个不同的走子法以将死对方. 当然，白方会走那一步棋.）

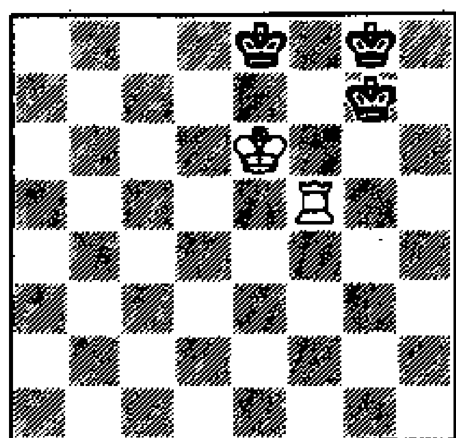


图 13

白方走棋.

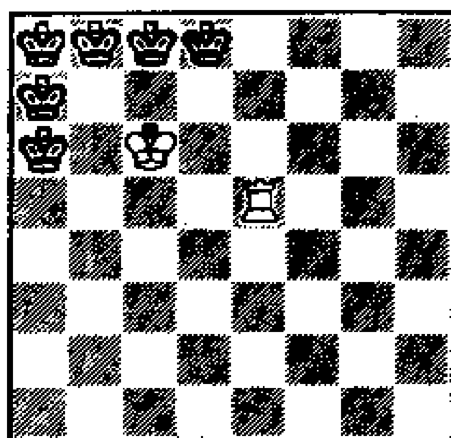


图 14

问题 2. 白方应如何尽快将死对方.

1c. 黑方王棋被限制于三行范围内. 当黑方王棋被限制在三行或三列时, 有一些办法可改进第1、2节中的弈法. 图13显示了白方走过R—f5+后的的局势. 黑方王棋在两个象限之一, 白方可避免将其棋子走到棋盘边上之烦. 他可以走: 1. R—f6, 2. K—e7. 如果K—e7是合法的, 则黑方王棋在右边的象限内; 若这一步是非法的, 则黑方王棋在左边象限内.

图14表明在某些局势下, 白方能节省时间的其他方法. 在1. R—d5+之后黑方王棋可逃到棋盘的另一边并在那里苟延残喘, 一直拖到白方的第十二步才完蛋. 试问, 白方应怎样走, 才能更好些? (答案见下文.)

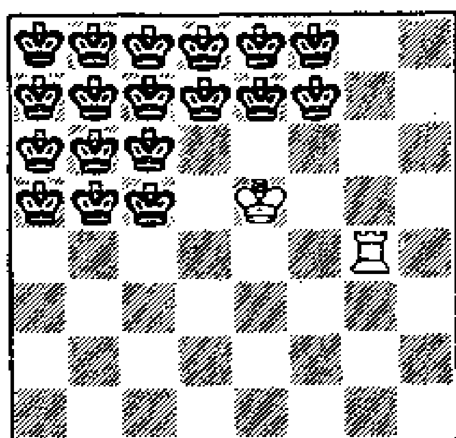


图 15

白方走棋.

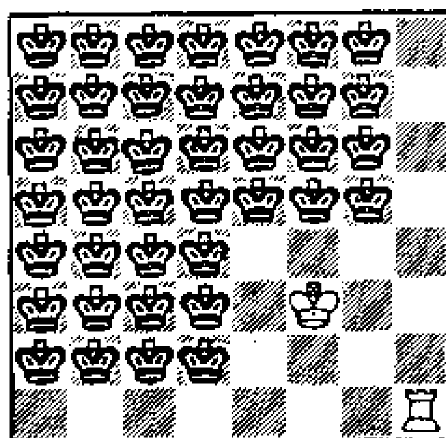


图 16

白方走棋.

1d. 黑之王棋限于四行. 图15类似于图13. 白方希望走R—f4. 如果他现在走这步棋, 而这是一着"将", 则黑方王棋将在两个象限之一, 白方如按图13的办法则是距边缘太远了. 白方仍可以走1. K—e6以避免第2节所讲方式中的虚掷时间与精力. 如果这着非法, 他可走…R—g5, 这时他已削减了较小的象限之大小. 如这一着是合法的, 他可走2. R—f4. 如果这着是"将", 他即可照图13后面的方式着下去.

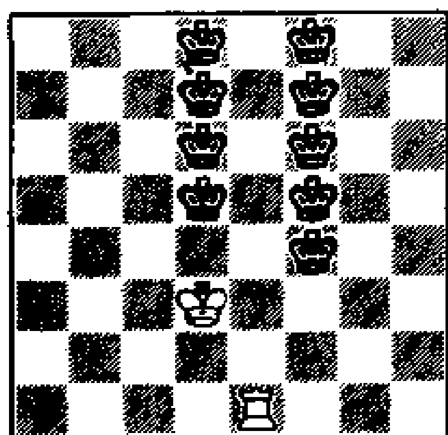


图 17
问题 3. 白方走棋.

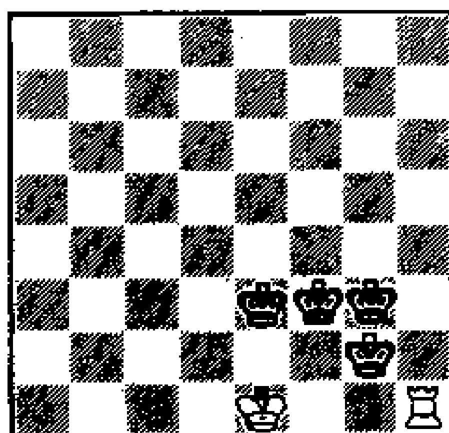


图 18
问题 4. 白方走棋.

1e. 黑方王棋被限定的范围超过四行. 当黑方王棋的活动范围是棋盘上很大一片区域时，白方企图尽快削减此区域. 图 16 表明黑方王棋被限定的最大可能的一片区域，如果区域足够大，白方能想办法削减黑方王棋所占据的较小的象限的尺寸(而不是像第 1 节所说的较大者). 由图 16，弈法可如下进行：1. R—g1，2. K—e3，3. R—f1. 不论移动堡垒的哪一步是"将"，白的下一步棋就是把他的堡垒向左走两格以造成两只王棋同在一象限，而最小一侧至多只有四格的局势. 图 17 是更为难下的，目标仍然是要把黑方王棋限制在至多四行之内，而不把王棋与堡垒像 2 那样走到棋盘的边上.

图 18 是由图 10 所产生. 如果 1. K—f2 是非法的，则白方应当怎样走棋，才能把黑方王棋限制在不超过四行之内？(注意一个 3×4 矩形包含了黑方王棋所可能占据的一切方格，但矩形的边并不由棋盘的边与堡垒来定出.)

本文业已证明，在"瞎子打仗"棋戏中，具有一只王棋与一只堡垒的白方一定能够将死只有王棋的黑方. 1a 与 1b 两节证明了，如果白方能把黑方王棋驱赶到棋盘边上，他就一定能够将死对方. 第 3 节表明白方能够轻易地达到一种局势，其时黑方王棋所可能占据的一切

位置是一个矩形，它由堡垒所控制的一行一列与棋盘的两边所构成.第1与第2两节已证明白方能从这样的一种棋局过渡到矩形更小的棋局.这里我们所说的矩形A小于矩形B的意思是指，A的较小尺寸小于B的较小尺寸，或者当A、B的较小尺寸相等时，A的较大尺寸小于B的较大尺寸.上述论证已足够证明白方能在一大批棋局状态中，以及在任意大的矩形棋盘上将死黑方.

在仔细研究了这里详加列举的战术以后，可以证明从图10的初始状态出发，白方至多用上39步即可将死对方，于是也就用不着再去担心"五十步不赢就算和棋"的规则了.

致　　谢

本文是斯坦福大学CS204班级在1978年秋季教学中一个指定的课外作业的结果.我要向任课教授唐纳德·E·克努特(Donald E. Knuth)先生表示谢意，他从德国友人处听到这种游戏后设计了这个题目.我还要感谢本课程的助教克里斯·凡·维克(Chris Van Wyk)，他对问题作了详尽充分的讨论并帮助笔者编写了本文，CS204班上的学生也提出了一些有价值的建议.

解　　答

1. 三步棋即可救出堡垒.R—f1，K—f1与K—f2.在后两步棋的每一步下过之后，黑方都不能吃掉堡垒，因若如此的话，白方的走法就将被认为是非法的.如果走法是非法的，白方就当然可以随他高兴地使其堡垒远离虎口.

2. 白方用下述走法可以将死黑棋：1. K—c7，R—e6(如K—c7合法，白方可以用类似方式获胜，但着法将短得多).2. K—d7，K—d6，3. K—c7然后像1b那样再走9步便可将死.

3. 白方用下述手段把黑棋限制在更"小"的矩形之内：1. R—

e4,R—c4. 如果第一步是“将”,白方可应之以 2. R—e2,如果第二步是“将”,白方可续走 3. R—e4.

4. 白方首先将堡垒移到安全地带. 1. R—h5, 2. R—a5. 然后白将王棋移近堡垒 3. K—d2, 4. K—c3,如果王棋的第二次移动为非法,他应把堡垒再移到安全地带 R—g5,并继续用下列走法试图把其王棋移近堡垒,即

5. K—e3,K—c3,6. R—a5,如果这两步王棋走法均属非法,则黑方王棋必在 d4 位置. 白方可走一着 6. R—h5,使王棋移动并再继续试探. 这一次,他必定取得成功. 如果 3. K—d2 是非法的,则白方可走 R—a4 以及 4. R—h4 以产生一种局势,它将会类似于这步棋为合法时所造成的形势.

参考文献

1. Compayne, Charles. 1976. Kriegspiel. *Games and Puzzles* 50: 12—15.
2. Fine, Reuben. 1941. *Basic Chess Endings*. David McKay.

心理扑克

● 麻省理工学院

□ 艾迪・夏米尔(Adi Shamir)

□ 劳纳德・L・里凡斯特(Ronald L. Rivest)

□ 利昂纳德・M・阿德曼(Leonard M. Adleman)❶

本文提要

两位不太诚实的对手能否不用真的扑克牌——例如,改用打电话方式——来进行一场公平的扑克游戏?本文将提供两个截然不同的答案:

1. 不行.(附以严格的数学证明.)

2. 行.(给出正确与完整的谈话记录.)



有一天,两位"心理象棋"专家对这种游戏感到了厌烦.于是其中一位说道:"让我们变变花样,来做个'心理扑克'游戏吧.""好啊,让我来发牌!"另一个人说.



---

❶ 译者注:因设计不能破译的数论密码而近年来声誉鹊起、名闻遐迩的三位专家.

我们的故事与下述问题有关(由罗伯特·W·弗洛德(Robert W. Floyd)提出):"是否有可能做一种公正的'心理扑克'游戏?"我们将对此给出一个完整而又令人迷惑不解的答案. 首先我们将证明问题本质上无解,而后又描述一种玩"心理扑克"的公正办法.

玩"心理扑克"是什么意思?

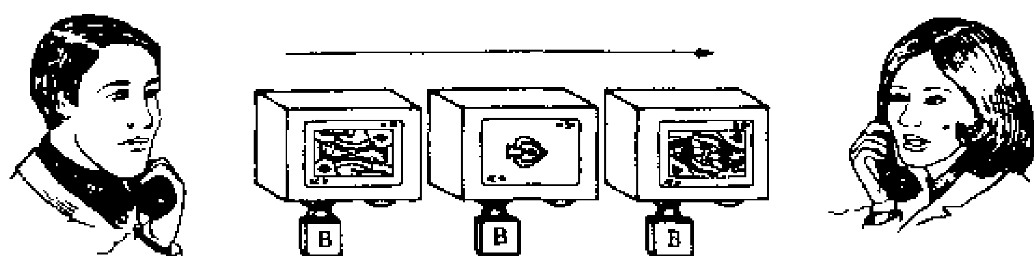
"心理扑克"游戏的玩法同普通扑克(参看 Hoyle〔2〕)完全一样,不过并没有一张牌作为道具,局中人之间的一切联系都得用传递信息的方式来进行. 如果我们设想有两个局中人,鲍勃(Bob)与爱丽丝(Alice),想用打电话的方式来做游戏,也许有助于阐明"心理扑克"的基本规则. 由于不可能把扑克牌在电话线上传来传去,整个游戏(包括发牌在内)必须自始至终地通过两人的口头语言(或者用数字传递办法)来实现.

我们假定两人都不老实. 即使他手中根本没有爱司,他也可以轻易说上一句"有一张爱司露头了"!不过,我们所说的"心理扑克"的公正玩法必需排除任何性质的欺诈行为,否则就不能称为"公正"了.

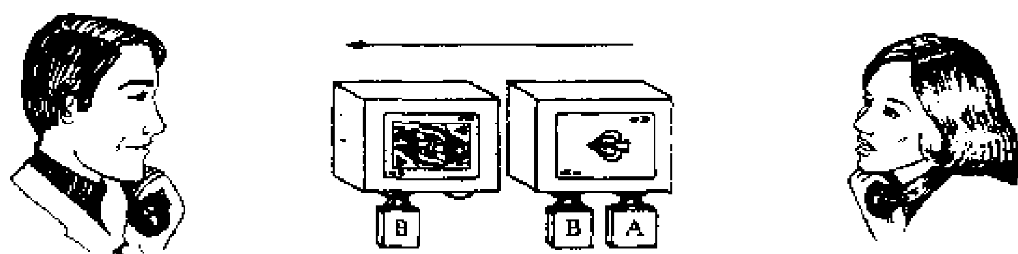
公正的玩法必须从"公正的发牌"开始. 为了做到这点要求,局中人可按某些事先商定的步骤来交换一系列信息.(这些步骤有可能要求双方利用骰子或其他随机数发生装置以估量他手中的牌或他所传输的信息.)每位局中人都必须知道他手上有些什么牌,但又必须对别人手上的牌一无所知. 发牌方法必须保证双方手上的牌完全独立,毫不相关,而且对任何游戏者来说,各种牌在他手上出现的可能性应当完全一样.

在游戏进行过程中,局中人可能要求从剩余的一堆牌里再抽取一些新牌,或者要求向对方亮出他们手中的某几张牌. 他们应能做到此点而不至于泄露手上其余牌的机密.

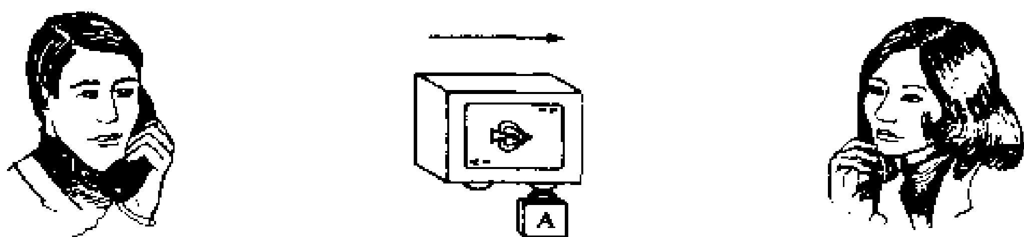
在游戏的结尾,每一方都要有可能进行核查,以保证该游戏确是公正无欺地在进行,对方并无欺诈行为. 如果一方宣称他发到了四张



鲍勃对扑克牌作了"加密",将它打乱后送给爱丽丝.



爱丽丝选了一张牌给鲍勃,另一张给她自己,并对之进行了"加密",然后把它们一起送还给鲍勃.



鲍勃对两张牌都作了"解密",并把爱丽丝加密过的一张牌退还给她.

图 1

爱司,另一方必须有办法对之进行核实.

使"心理扑克"成为一种公正游戏的上述一系列要求看来似乎很难得到满足.为了把事情搞得简单一些,我们可以假定双方都有计算机,这样就有可能采用非常复杂的谈话记录(例如,也包括加密过程).但我们并未假定任何一方会信赖别人的计算机(双方都可以对其计算机编制程序,教它们进行欺骗).

我们建议你在阅读下文之前，试图找到一个你自己的玩"心理扑克"的办法，你一定会觉得那是一种非常有趣的挑战.

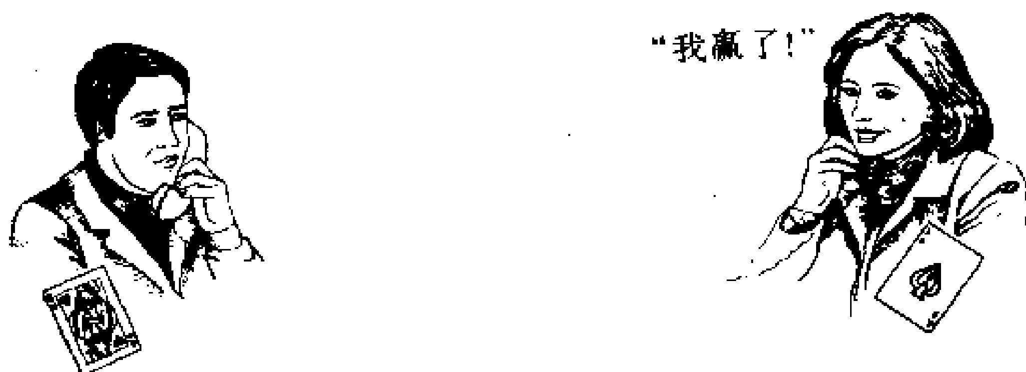


图 2

爱丽丝"解密"了她的牌，然后他们互相比较牌的点数大小以决定谁是胜者.

结　　论

我们将对"心理扑克"问题提供两种截然不同的回答：

1. 在理论上提供一个严格的证明，即任何一种发牌方式都不可能同时满足下列要求：双方手中的牌毫无关联，一方对另一方手中的牌一无所知(除了他知道对方手中的牌与他不一样之外).

2. 存在一种巧妙的谈话方式可用于发牌，使人们有可能按照规定要求来做公正无欺的"心理扑克"游戏.

两种答案的公然矛盾并非由于它们中间出了什么毛病或失误. 事实上，我们倒是想把一项十分有趣的任务留给读者，让他们自己去解释：之所以导致这种明显矛盾的原因究竟是什么？要推敲出其中的微妙差别.

不可能性的证明

为了简单起见，让我们考虑最低限度的非平凡情况，即从三张牌的集合$\{X,Y,Z\}$里分发两张不同的牌(每一方发到一张牌). 此种情

况下的不可能性证明可以极容易地推广到牌的任何组合，以及各人手中牌数不止一张的情况.

如果这一情况下，合乎要求的谈话记录能够存在，则在交换过一系列有限信息之后，爱丽丝与鲍勃都知道了自己手里的牌而不知道别人的牌. 当然这些信息与两位局中人的选牌必须协调一致，以防止两人得到同样的牌.

假定在一次特殊的发牌中，

交换的信息是 $M_1, \cdots, M_n$，

爱丽丝实际拿到的牌是 X，

鲍勃实际拿到的牌是 Y.

我们把 S_A 定义为爱丽丝在具有相同交换信息的任何发牌中所能拿到的牌的集合（由于每位局中人都希望有随机选择的自由，以防别人猜到他的牌，于是产生了具有同样交换信息的不同发牌方式）. 显然，牌 X 是在集合 S_A 里的.

如果 S_A 只有这张 X 牌，则此种发牌就与鲍勃对爱丽丝的牌应该一无所知的要求发生了抵触. 显然，在此种情况下，信息序列唯一决定了爱丽丝的牌，所以，从信息论的观点来看，鲍勃对她的牌已有了（完整的）信息. 进一步说，在任何一种物理上可以实现（而且有尽的）的发牌谈话记录中，爱丽丝只可能有有限多种随机计算的自由，因此鲍勃只要检查一下其中有哪些可以与给定的信息序列相容，就能实际上查出爱丽丝手中的牌.

另一方面，如果 S_A 包括了全部三张牌，则鲍勃就无法拿到牌，因为不论他拿到哪一张牌，信息序列将容许爱丽丝也能拿到同样的牌. 于是，S_A 只能含有两张牌.

鲍勃能拿到、而不致于改变其外部性态的牌的集合 S_B 也可类似地定义，因而，这集合也必然含有两张牌. 可是，牌的总数只有三张，因此 S_A 与 S_B 是不可能没有公共元素的（在我们这个例子中，牌 Z 同属于两个集合）. 因此，在信息序列 $M_1, \cdots, M_n$ 的情况下就会发生鲍勃与爱丽丝同样拿到 Z 牌之事. 所以，谈话记录无法保证鲍勃与

爱丽丝能选取不同的牌．于是，我们可以下结论了：公正的发牌是不可能的．

可供发牌用的一个谈话记录

下列解法可以满足问题的一切要求．首先，鲍勃与爱丽丝同意采用一对加密与解密函数 E 和 D，它们具有下列性质：

1．$E_K(X)$是信息 X 在采用密钥 K 时的加密形式，

2．对一切信息 X 与密钥 K 都有

$$D_K(E_K(X))=X,$$

3．对一切信息 X 与密钥 J,K 都有

$$E_K(E_J(X))=E_J(E_K(X)),$$

4．给定 X 与 $E_K(X)$后，密码分析家将无法通过计算来推出密钥 K，对一切 X 与 K 都是如此，

5．给定信息 X 与 Y 后，无法通过计算找到密钥 J 与 K，以使得 $E_J(X)=E_K(Y)$．

上述第 3 条性质，即加密的可交换性显得有点不寻常，但不是不可能做到的．性质 4 与 5（特别是性质 4）则实质表明 E 的保密性极强，或者说，它是不可破译的．

作为具备以上性质的函数实例，可以考虑

$$E_K(M)\equiv M^K(\bmod n),$$

这里 n 是一个很大的数（素数或是有给定因子的合数），而且为鲍勃与爱丽丝所已知，此外还有 $gcd[K,\varphi(n)]=1$，

这里的$[\varphi(n)]$是欧拉数论函数，从数 n 的质因子分解式中极易算出它来．

与此相应的解密函数是

$$D_K(c)\equiv C^L(\bmod n),$$

此处

$$L\cdot K\equiv 1[\bmod\varphi(n)].$$

由于

$$E_K[E_J(M)] = E_J[E_K(M)] \equiv M^{JK} (\mathrm{mod}\ n),$$

因而 E 能满足性质 3.(有关这一函数的保密力量及重要性的更多细节,请参阅文献[1,3,4].)我们之所以在这里描述这个特殊加密函数,仅仅是想说明我们所需要的加密函数确实是存在的,除了性质 1 至 5 以外,我们不准备利用这个函数的任何其他特殊性质.

一旦鲍勃与爱丽丝同意采用函数 E 与 D(对我们这个例子来说,这意味着他们同意使用 p),他们就分别选择了机密的加密密钥 B 与 A. 这些密钥将保密到底,直至游戏结束. 其时,他们可以把密钥公开曝光,以验证在整个过程中并无欺诈行为.

鲍勃现在可拿起 52 个信息:

"梅花 2",

"梅花 3",

…………

"黑桃爱司",

并将每个信息(它的二进位字符行可看成一个数)用他的密钥 B 来进行加密.(这就是说他需要计算 E_B("梅花 2"),等等.)然后,他把已经加密的一叠牌进行洗牌(随机地重排),并把它们统统电传给爱丽丝.

爱丽丝随便选择了五张牌(信息)并把它们送还给鲍勃. 鲍勃通过解密,了解到他手中有些什么牌. 但爱丽丝根本不知道鲍勃手中有些什么牌,因为用于编码的密钥 B 只有鲍勃才知晓.

现在爱丽丝选取五个其他信息,用她的密钥 A 进行加密,然后把它们送给鲍勃. 现在,五个信息中的任一个都已被双重加密为 $E_A(E_B(M))$或者与之等价的 $E_B(E_A(M))$,对每个 M 都是这样. 鲍勃解密了这些信息,对五个信息都作出了 $E_A(M)$,并把它们送还给爱丽丝. 爱丽丝可以利用其密钥对之解密,这样就得出了她手中的牌. 由于鲍勃不知道 A,所以他对爱丽丝的一手牌毫无所知.

迈克尔·拉宾(Michael Rabin)对上述过程建议提供一种物理模

拟.我们可以把加密看作是在存放扑克牌的匣子上装一把挂锁.鲍勃先把所有扑克牌装在外表全无区别的一些匣子里,再在匣子外面装上挂锁(它们的开锁钥匙全是 B).爱丽丝挑选了五只匣子还给他以作为他的一手牌,另外又外加五只匣子,这些匣子的外面又给扣上了她的挂锁(开锁钥匙为 A).接着,鲍勃把十只匣子上他的挂锁统统拿掉了,然后把匣子上仍然装有爱丽丝挂锁的东西退还给她,以作为她的一手牌.请注意,在上面的叙述中,已默认了在装锁与开锁的先后顺序上的可交换性.

在游戏进行过程中不论哪一方需要添牌时,则可对每一张牌重新执行上述的步骤.

游戏结束时双方都亮出了他们的密钥.现在,每个局中人都能核实,他的对手所声称之牌,确实是真正发到他手里的.根据性质 5,任何一方都不能利用一个不是他真正用过的、而是另外一个密钥(后者也许能使他拿到较好的一副牌)来进行欺骗.

上述过程也可以推广到局中人超过两个的场合.(我们把细节留给读者自己去思考.)另一种明显推广是把可交换的编码函数应用于秘密通讯系统,以传送任何信息(而不是无足轻重的扑克牌之类东西),即使通讯渠道里有人在偷听也不要紧.

结　论

一开始我们证明了发牌问题是无解的,可是接着又对该问题提出了一种实际上管用的解法.现在我们要把这个哑谜留给读者.请问,怎样才能自圆其说?(提示:每位局中人按理说,都能从现有信息中完全决定出对方手中所有的牌,如果不是由于"破译"密码所需花出的巨大计算工作量使他只能望洋兴叹的话.)

致 谢

我们应当向罗伯特·W·弗洛德、迈克尔·拉宾以及阿尔伯特·迈耶(Albert Meyer)表示谢意,感谢他们的热心推动与有价值的建议.

参考文献

1 Diffie, Whitfield and Hellman, Martin E. 1976. New Directions in Cryptography. *IEEETrans. Info. Theory* IT-22:644-654.

2 Morehead, A. H., Frey, R. L. and Mott-Smith, G. 1947. *The New Complete Hoyle. New York: Garden City Books.*

3 Pohlig, Stephen C. and Hellman, Martin E. 1978. An Improved Algorithm for Computing Logarithms over GF(p) and its Cryptographic Significance. *IEEE Trans. Info. Theory* IT-24: 106-110.

4 Rivest, Ronald L., Shamir, Adi and Adleman, Leonard M. 1978. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *CACM* 21: 120-126.

原注: 本研究由 NSF 拨款项目 MCS78—05849, MCS78-04343 以及 ONR 拨款项目 N00014—76—C-0366 提供资助.



● 麦克吉尔大学

□ 瓦赛克·克瓦塔(Vašek Chvátal)

在 R. C. 麦克拉根(R. C. MacLagan)编著，于 1901 年由戴维·纳脱(David Nutt)在伦敦出版的一本书《阿格莱郡的游戏与消遣》的第 128 页上，人们找到了一个古代苏格兰游戏的如下描述.

便宜、适中、高价

它也是个两人游戏. 字母 C、M、D 分别表示 Cheap、Middling、Dear 的缩写，它们被分别书写在石板上，中间适当地留着些间隔. 在字母 C 的下面有三个数字 1,2,3；M 的下面写着 4,5,6；D 的下面写着 7,8,9，就像下面的样子：

C	M	D
1,2,3	4,5,6	7,8,9

先走的 A 从任意一组中认定一个数字，但要瞒住 B，不能让他知道. 接着，就开始猜谜式的一问一答. A 先说："我爸爸从市场里买回来一匹马."B 问道："便宜、适中、还是出了高价?"A 于是作出回答，把他所选定的数字所属的那一组告诉给 B，譬如说，如果他选中的数字是 5，他的回答就一定要说是"代价适中". 然后，B 就从三个数字中猜一个. 如果他恰巧猜对 5，他就得了 5 分，但若他猜 4 或 6，

则 5 分就归入 A 的名下. 不论属于哪种情况, 数字 5 就要从石板上涂抹掉. 接下来轮到 B 选数, 让 A 去猜, 双方交替地进行, 直到全部数字勾销为止. 最后把双方名下的分数统统加起来, 谁得的分数较高, 就算他赢.

下面我们将要说明这种"便宜、适中、高价"游戏的分析结果, 为了使不熟悉博弈论的读者易于读懂, 让我们先把有关矩阵博弈的一些基本事实简单地勾划一下.

让我们从一个猜辅币的游戏开始, 它和"便宜、适中、高价"游戏有些瓜葛, 可是要简单得多. 该游戏由两人对垒, 每人手中都藏起一枚五分镍币, 一枚一角银币或一枚二角五分辅币. 然后, 双方同时拿出一枚来亮相, 如果它们属于同一品种, 则就归第一位局中人所有. 否则, 它们就属于第二位局中人. 很明显, 无论哪一方都不具有保证稳赢的策略: 如果运气不佳的话, 有可能每次都输. 可是我们仍然可以认为, 至少就统计意义来说, 第二位局中人处于较为有利的地位. 如果他正确地做游戏, 他可以期望在长期赌赛中取得稳定的收入. 自然, 他必须用某种随机的方式作选择, 以使对方不能预料. 但与此同时, 各种辅币的出现概率必须予以固定. 这就是说, 在长期赌赛时, 第二位局中人必须用全部时间的$\frac{7}{34}$出示镍币, $\frac{12}{34}$的时间出示银角子, 而以余下来的$\frac{15}{34}$时间亮出二角五分辅币. 假定他确实完全按照上述指示行事, 我们现在来看看长时间反复赌赛的情况. 第一位局中人所出的每一枚镍币, 将有$\frac{7}{34}$的时间遇上第二位局中人所出的镍币, $\frac{12}{34}$的时间遇上一角银币, $\frac{15}{34}$的时间遇上二角五分币. 不论第一位局中人如何行动, 他不能忽视由第二位局中人的随机选择而引入博弈的机遇因素. 在上述这群赌赛中, 第二位局中人将有$\frac{7}{34}$的时间输掉五分镍币, 而将有$\frac{27}{34}$的时间赢进五分镍币. 类似地, 在第一位局中人出示一角银

币的那些赌赛中，第二位局中人将有$\frac{12}{34}$时间输掉一角，而在$\frac{22}{34}$的时间中赢进一角.在第一人选择二角五分辅币的赌赛中，第二人将有$\frac{15}{34}$时间输掉二角五分，而有$\frac{19}{34}$时间赢进二角五分.所以，总的看来，第二位局中人在每次赌赛中平均赢$\frac{100}{34}$美分.当然，这种收益并不是保证可以拿到的，它不过是一种期望值，这正像抛掷一枚没有偏差的钱币，出现正面的机会是百分之五十只是一种期望而不是保证一样.从这种意义说，第一位局中人可以保证他的每局损失不至于超过$\frac{100}{34}$美分，他只要采用如下的混合策略：$\frac{10}{17}$时间出镍币，$\frac{5}{17}$时间出银角子，余下$\frac{2}{17}$时间出二角五分辅币.现在不论第二位局中人如何行动，每次他出镍币、出银角子，或者出二角五分，他的最大期望收益分别如下：

$$-\frac{10}{17}\cdot 5+\frac{5}{17}\cdot 10+\frac{2}{17}\cdot 25=\frac{50}{17},$$

$$\frac{10}{17}\cdot 5-\frac{5}{17}\cdot 10+\frac{2}{17}\cdot 25=\frac{50}{17},$$

$$\frac{10}{17}\cdot 5+\frac{5}{17}\cdot 10-\frac{2}{17}\cdot 25=\frac{50}{17}.$$

这个博弈可用矩阵表示如下：

$$\begin{bmatrix} 5 & -5 & -5 \\ -10 & 10 & -10 \\ -25 & -25 & 25 \end{bmatrix}.$$

第一位局中人从三行中选一行(分别相当于出示镍币、银币与二角五分币)，不知对方行动的第二位局中人则从三个列中挑选一列.行列交叉处的矩阵元素则为第一位局中人的所得金额.一般地说，每一个矩阵确定了一种博弈，一方选取行，另一方选取列，相应的 a_{ij} 则表明第一位局中人所获得之收益.(“便宜、适中、高价”游戏也可看作这种游戏，矩阵的每一行相当于局中人 A 在全部九轮猜数过程中所

能采取的明确的策略. 类似地, 每一列相应于 B 所能用的策略. 这个矩阵的行数与列数是很庞大的.) 其次, 每位局中人唯一有意义的策略是他的选择必须随机, 使他的对方猜不透. 假定第一位局中人以相对频数 x_i 来选取各行 $i=1,2,\cdots,m$, 第二位局中人以相对频数 y_j 选取各列 $j=1,2,\cdots,n$. 则第一位局中人的平均获得是(对每一局赌赛来说):

$$\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j.$$

如采用矩阵记法, 这一数量可记为 xAy. 分量为 $x_1,x_2,\cdots,x_m$ 的行向量 x 与分量为 $y_1,y_2,\cdots,y_n$ 的列向量 y 都具有共同性质: 所有分量均为非负数, 而且其和为 1. 这种向量称为*随机的*. 本世纪二十年代后期, 约翰·冯·诺伊曼(John von Neumann)在其论文《公司博弈理论》(1928 年, 载于《数学年刊》第 100 卷 195－320 页)中证明了著名的最小最大定理, 断言

$$\max_x \min_y xAy=\min_y \max_x xAy$$

对一个任意矩阵 A, 在所有的随机向量 x 上取最大值, 所有的随机向量 y 上取最小值时, 此式都成立. 如果用另一种方式叙述, 则对任一矩阵 A, 必存在随机向量 x^* 与 y^* 以使得

$$xAy^* \leqslant x^*Ay^* \leqslant x^*Ay$$

对一切随机向量 x 与 y 都成立. 现在 x^* 与 y^* 可看作是由 A 所表示的博弈的最优策略: 第一位局中人可以期望每局至少获得

$$\min_y x^*Ay=x^*Ay^*.$$

而第二位局中人则可指望, 每局支付不大于

$$\max_x xAy^*=x^*Ay^*.$$

数量 x^*Ay^* 被称为博弈的值. (上述猜辅币游戏的值是－50/17.)

讲过这些预备知识之后, 我们回到"便宜、适中、高价"这个本题. 为了打破悬案, 让我们来揭示一下, 就此博弈而言, B 较为有利; 博弈的值大约是－3.8(这就是说, 如果双方都按最优策略行事, 则 B 可

以指望获得 24.4 分，而 A 的分数只有 20.6 分）. 游戏中任一可以想象得到的状态都被留在石板上的数集 S 所明确规定出来. 如 S 中数的个数为奇数，就轮到 A 选择一数；如为偶数，则由 B 挑选一数. 如果挑选数字的局中人按照他的最优策略办事，则在余下的游戏中，不管对手采取什么策略，他都有希望获得 v 分. 与此类似，倘若猜数的局中人遵循他的最优策略，则不管他的对手采用什么办法，在余下的游戏中，他都有希望获得 w 分. 由最小最大定理可知，和 $v+w$ 应等于 S 中所有数字之和. 我们将把差 $v-w$ 看作 S 的值. 我们将要叙述一个简易方法，以计算每个 S 的值，还要说明从 S 中挑选一个数的最好办法，以及在 S 的一个指定分组中猜测一个数的最好办法. 我们的主张可通过对 S 的大小施行归纳法而得到证实，不过此处略去了烦琐的证明过程.

游戏的每一轮开始时，一位局中人提出他已认定数字的一组数. 最优策略如下：

1. 较大的组总是优于较小的组；

2. 在三个数字所成的一组中，高价优先于适中，适中则优先于便宜；

3. 所有的二个数的组都是同样诱人的；

4. 对只有一个数的组来说，便宜要比适中优先，而适中则优先于高价.

当然，数的选取必须随机化. 策略如下：

5. 在包含二个数 x 与 y 的一组中，

要以概率 $y/(x+y)$ 选取 x；

以概率 $x/(x+y)$ 选取 y.

6. 在包含三个数 x、y、z 的一组中，

要以概率 $yz/(xy+xz+yz)$ 选取 x；

以概率 $xz/(xy+xz+yz)$ 选取 y；

以概率 $xy/(xy+xz+yz)$ 选取 z.

对选数的局中人来说，上述规则提供了一个最优策略的完整说

明. 事实上，几乎所有的最佳选择都具有上述形状，唯一的例外是

7. 当"便宜货"只有一个数，"中档货"两个数，"高价品"两个数时，则玩"便宜货"这一招决不会错.

现在我们要转向各种状态的计值问题. 首先让我们假定，三组中的两组已从黑板上完全抹掉. 在此种情况下，$\{x\}$的值是$-x$，$\{x,y\}$的值是$(x^2+y^2)/(x+y)$，而$\{x,y,z\}$的值是

$$\frac{1}{xy+xz+yz}\left(xyz-\frac{xy(x^2-y^2)}{x+y}-\frac{xz(x^2+z^2)}{x+z}-\frac{yz(y^2+z^2)}{y+z}\right).$$

直截了当地说，最后这个公式告诉我们以下结果：

$\{1,2,3\}$的值是$-\frac{613}{330}\doteq-1.85757575$，

$\{4,5,6\}$的值是$-\frac{64913}{18315}\doteq-3.544253344$，

$\{7,8,9\}$的值是$-\frac{2129473}{389640}\doteq-5.465232009$.

接着，我们注意到，把上述1—6条规则与出现的21个值相结合，即能指出一种简单办法来计算一些不同状态的值. 例如，我们来考察$S=\{1,2,3,4,5,7,9\}$. 由于S的元素个数是奇数，故轮到A来选数. 一种较佳的游戏进程可能取如下方式：

A选便宜， B选便宜， A选中档， B选高价， A选便宜，B选中档， A选高价.

两位局中人也有可能按下列方式做游戏：

A选便宜， B选便宜， A选便宜， B选中档， A选中档，B选高价， A选高价.

因而我们可以下结论说，S的值是

$\{1,2,3\}$的值$+\{4,5\}$的值$-\{7,9\}$的值

$$=-\frac{613}{330}+\frac{41}{9}-\frac{130}{16}=-\frac{21491}{3960}\doteq-5.4270202.$$

为了给512个状态的每个状态提供一个类似的简便计值方法，我们将需要一些另外的补充材料. 再一次设S为三组中的两组已完全抹掉的状态. 我们现在要介绍在$S=\{x,y\}$上进行的出格游戏，其

玩法是两轮均由 A 来选数；至于在 $S=\{x,y,z\}$ 上玩的出格游戏，其意思是 A 仅仅在第一轮中选数，而在余下的两轮中猜数. $\{x,y\}$ 的出格值是 $-(x^2+y^2)/(x+y)$，而 $\{x,y,z\}$ 的出格值是

$$\frac{1}{xy+xz+yz}\left(xyz+\frac{xy(x^2+y^2)}{x+y}+\frac{xz(x^2+z^2)}{x+z}+\frac{yz(y^2+z^2)}{y+z}\right).$$

直截了当地说，最后一式告诉我们

$\{1,2,3\}$ 的出格值是 $\frac{973}{330}\doteq 2.948484848$，

$\{4,5,6\}$ 的出格值是 $\frac{124313}{18315}\doteq 6.787496587$，

$\{7,8,9\}$ 的出格值是 $\frac{4185793}{389640}\doteq 10.74271892$.

现在我们业已具备对任一状态快速求值的一切必需知识. 例如，我们来考虑 $S=\{1,2,3,5,6\}$. 一种较好的游戏进程如下：

A 选便宜， B 选便宜， A 选中档， B 选便宜， A 选中档，

两位局中人也可同意照如下方式进行：

A 选便宜， B 选便宜， A 选便宜， B 选中档， A 选中档.

这样，我们断言，S 的值可按如下方式计算：

$\{1,2,3\}$ 的出格值 $+\{5,6\}$ 的出格值

$$=\frac{973}{330}-\frac{61}{11}=-\frac{857}{330}\doteq -2.5969696.$$

最后，我们转向最佳猜数策略：假定一位局中人要从一个三数组 $\{x,y,z\}$ 挑选一数. 令 a,b,c 分别表示 $S-\{x\}$，$S-\{y\}$，$S-\{z\}$ 的值. 显然，另一位局中人的最佳猜数策略与由下列矩阵

$$\begin{bmatrix} -x-a & x-a & x-a \\ y-b & -y-b & y-b \\ z-c & z-c & -z-c \end{bmatrix}$$

所决定的博弈中，第二人的最优策略是一致的.

所以这位局中人应当

以概率 $\frac{1}{2(xy+xz+yz)}[y(x+c-a)+z(x+b-a)]$ 猜数 x；

以概率$\frac{1}{2(xy+xz+yz)}[x(y+c-b)+z(y+a-b)]$猜数 y；

以概率$\frac{1}{2(xy+xz+yz)}[x(z+b-c)+y(z+a-c)]$猜数 z.

类似地，如果一位局中人从一个二个数的组中选出一数，而 a,b 分别为 $S-\{x\},S-\{y\}$的值，则另一位局中人的最佳猜数策略相当于第二位局中人在下列矩阵博弈

$$\begin{bmatrix} -x-a & x-a \\ y-b & -y-b \end{bmatrix}$$

中的最优策略. 这就是说，该局中人应以

概率$\frac{1}{2(x+y)}(x+y-a+b)$猜数 x，

而以概率$\frac{1}{2(x+y)}(x+y+a-b)$猜数 y.

原注：此文引自技术报告 SOCS—79.3.

随机独脚跳问题,怎样使孩子多读一些

● 纽约州立大学宾厄姆顿分校

□ 戴维·贝伦古特(David Berengut)

有不少深孚众望的游戏把随机因素(如掷骰子,转螺旋,或分发一叠已洗过的扑克牌)同一个游戏场地或特殊棋盘巧妙地结合了起来.通常,这种棋盘上画着一系列空格,有的构成封闭回路(如在"垄断资本家"游戏中),有的则具备各自分开的起讫点(如在纸牌或十五子游戏中).在这些游戏中,可以想入非非地把棋子的移动看作一种随机性的小孩独脚跳.作为一个数理统计学家,我对此类游戏的概率特性有着相当浓厚的兴趣,远不止只有五分钟热度.

这篇文章起源于一个戴维·克拉纳(David Klarner)提示给我的问题.一位低年级教师设计了一种随机性的小孩独脚跳游戏,旨在提高她班上学生的阅读兴趣.棋盘上画着一系列空格,并有各自分开的起讫点.在一些格子中载有指令,要求学生去做各式各样的阅读练习,其他格子则完全是空白.每位学生都要在棋盘上轮流地行走,他们的移动步数则完全取决于一枚经过"修正"的骰子,其表面只有两个1点,两个2点和两个3点(因此移动的格子数只能是1格,2格或3格,它们的出现都是等可能的).如果学生正好停留在载有指令的格子上,他就一定要去做合适的阅读练习.

由于引进了机遇因素(怎样才能控制掷骰子的结果,同样是一个饶有兴味的挑战),这位教师极其成功地抓住了学生的注意力.教师

提出的问题如下：为了使她的学生阅读的期望数最多，她应当在哪些格子上放入指令？在所提出的这个问题中，当然蕴含着一个假设，即棋盘上的格子数以及阅读作业的数量都是固定不变的.

从语言过渡到数

解决许多数学问题的钥匙在于简单而正确地把问题算式化. 本问题在很大程度上也是如此. 由于在一般情况下求解数学问题通常要比特殊情况更为容易一些，所以我要让棋盘上的格子数与阅读练习的个数都取成任意数，并用字母 n 与 t 分别代表这两个数量.

在解决任何问题时，首先我们必须确切知道问题究竟是说些什么东西. 那么，所谓"要加以实施的阅读作业的期望数"究竟意味着什么？随机变量的期望值在概率论中是一个已经标准化的概念——要而言之，它是随机变量各个可取值的概率加权平均数. 如果 T 表示棋盘上一次假想运行中所要做的作业数，则 T 是一个随机变量，可以取 0 到 t 之间的任何整数值，并以一定的概率(目前尚未确定出来)取其中的每一个值. 令 $P(j)$ 表示 T 取 j 值的概率，这里 j 是 0 与 t 之间的任一整数，则 T 的期望值可记为 $E(T)$，由下式给出：

$$E(T)=0\cdot P(0)+1\cdot P(1)+\cdots+(t-1)\cdot P(t-1)+t\cdot P(t).$$

从起点开始，依照自然数顺序来标记棋盘上的格子是比较方便的，即在紧接着起点的第一格记作 1，第 2 格为 2，以此类推，直到最后一格记作 n. 令 $i_1,i_2,\cdots,i_t$ 按照递升顺序，表示其中载有作业指令的格子番号. 这样一来，对此种游戏的任意一局而言，T 的值就是在集合 $\{i_1,i_2,\cdots,i_t\}$ 中游戏者所曾驻足过的格子数. 这里，一种方便的数学手段是利用所谓**指示变量**，按照某一特定事件的发生与否，它只能取 1 或 0 值. 对本问题来说，如果 I_1 是游戏者驻足于格子 i_1 这一事件的指示变量，则当该游戏者在游戏进行过程中确曾驻足在格子 i_1 上时，I_1 就等于 1，如果他并未停留在此格子上，则 I_1 等于 0. 按照同样的方式，可令 I_2 为游戏者驻足于格子 i_2 这一事件的指示变量，

……以此类推，直至 I_t 为止. 于是，很明显，$I_1+I_2+\cdots+I_t$ 即可简单而明确地表示在游戏过程中，游戏者曾经停留于含有作业指令的格子上的次数，我们把这个数称作 T.

于是，计算 T 的期望值就相当于计算 $I_1+I_2+\cdots+I_t$ 的期望值. 在此场合，我们要利用数学期望的一项基本性质：和的期望值等于期望值之和.（例如：某地一年中雨天的期望数于该地一月份雨天的期望数，加上二月份的，三月份的，等等.）如用记号表示，则 $E(T)$ 等于 $E(I_1)+E(I_2)+\cdots+E(I_t)$.

那么，我们将如何计算 $E(I_1)$，能否举例说明一下？由期望的定义可知，$E(I_1)$ 可由 $0\cdot P(0)+1\cdot P(1)$ 给出，上式实际上就是 $P(1)$，$P(1)$ 是 I_1 取 1 这个值的概率，即游戏者停在格子 i_1 上的概率. 同样，$E(I_2)$ 即是游戏者停在格子 i_2 上的概率，等等. 如果我们以 p_i 表示游戏者停在格子 i 上的概率，此处 i 从 1 到 n，则我们可以写出下式

$$E(T)=p_{i1}+p_{i2}+\cdots+p_{it}.$$

于是，问题就变成怎样去求出 $p_1,p_2,\cdots,p_n$ 的值. 一旦有了这些数据，我们就可以解决那位教师的问题：合理地挑选布置作业的格子，以使得 $E(T)$ 最大. 显然，它们就是 t 个格子，其相应的 p 值是在所有的 $p_1,p_2,\cdots,p_n$ 中的 t 个最大者. 其结果是：教师应该选择 t 个格子，在其上布置作业，这些格子是游戏者最有可能在其上驻足的.

然而，要决定它们究竟是哪些格子，这可不是一个轻而易举的问题. 为了说明解法，我想先考察一个较为简单的问题.

一个较简单的问题

设想游戏规则稍有改变，在棋盘上的每步移动只能走一或二格，由抛掷一枚钱币来决定. 在这种情况下，我们能否算出在每一个格子上驻足的概率 $p_1,p_2,\cdots,p_n$ 呢？

自然，p_1 的计算是太容易了，它正好是第一步走一格的概率，即

$\frac{1}{2}$. 现在，p_2 是停留在第 2 格上的概率，这可以通过两种不同方式来达到：第一步走两格，或者第一步与第二步各自走一格. 第一种方式的发生概率是 $\frac{1}{2}$，而第二种方式的概率正好是两次连续抛掷钱币时出现两个反面的概率. 由于钱币的抛掷是相互独立事件，于是乘法原理告诉我们，这一概率不过是第一次出现反面的概率与第二次抛掷出现反面的概率之乘积，也就是 $\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$. 由于抵达第 2 号格子只有上述两种不同方式，因而，驻足于 2 号格子的总的概率 p_2 是以上两种概率之和，即 $\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$.

这种论证方式需要列举出所有可能的途径，自然也可用于计算 p_3 以及其余的各个 p 值. 但是，在棋盘上的格子逐渐向前推移时，计算工作量将变得沉重不堪. 幸而，这方面的困难可以通过如下办法来巧妙地回避. 那就是，按照一种递归方式来计算各个 p 值.

递 归 方 法

假定我们要去计算 p_i，即游戏者驻足于格子 i 上的概率，而 i 大于 2. 把棋子带到格子 i 上的任意一着必然来自第 $i-2$ 号或 $i-1$ 号格子. 因此，抵达格子 i 的两种不同方式可叙述如下：

1. 游戏者到达了第 $i-2$ 号格子，下一步移动 2 格（穿过了第 $i-1$ 格），

2. 游戏者到达第 $i-1$ 号格子，下一步移动是前进 1 格（见图 1）.

由于棋子的相继移动是独立事件，因而通过途径 1 驻足于 i 格的概率应等于停留在第 $i-2$ 号上格子的概率（p_{i-2}）与其下一步走 2 格的概率 $\frac{1}{2}$ 之乘积，即 $\frac{1}{2}p_{i-2}$；类似地，通过第 2 条途径到达的概率是 $\frac{1}{2}p_{i-1}$. 因此，停留在第 i 格上的总概率（p_i）等于 $\frac{1}{2}p_{i-1}+\frac{1}{2}p_{i-2}$. 由此

可知，在序列 $p_1, p_2, \cdots, p_n$ 中，从第三项起，每一项的值是其前面两项的平均数！

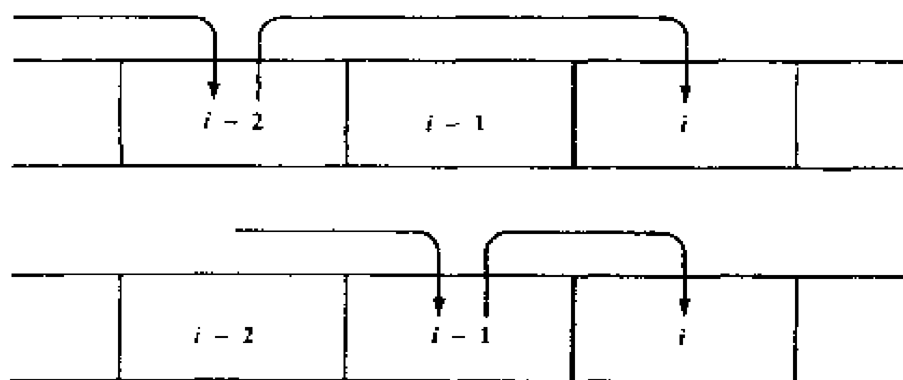


图 1

抵达第 i 号格子的两种不同方式.

已知 $p_1=\frac{1}{2}$，$p_2=\frac{3}{4}$，使我们能够断言 $p_3=\frac{1}{2}\left(\frac{1}{2}+\frac{3}{4}\right)$，即 $\frac{5}{8}$. 通过此种方法，我们能够递归地算出 $p_4\left(=\frac{11}{16}\right)$，$p_5\left(=\frac{11}{32}\right)$等等. 当然，如果我们有兴趣计算 p_{20} 的值，那就得把前面 19 个 p 值都算出来. 这对手算来说，未免太繁复了. 我们能否对任一 i，给出一个直接计算 p_i 的公式呢？答复是肯定可以的. 实际上，计算 p_i 的公式简单之至：

$$p_i=\frac{2}{3}+\frac{1}{3}\left(-\frac{1}{2}\right)^i.$$

$\left(\text{附带说一句，它将告诉我们 } p_{20}=\frac{2097152}{3145728}.\right)$

从递推关系式得到这个公式是一件相当简单的事情，我们首先观察到下述简单事实：两个数字的平均数必定介于这两数的正中间. 于是，p_i 在 p_{i-2} 与 p_{i-1} 的当中，因此，p_i 与 p_{i-1} 的差就绝对值而言，等于 p_{i-1} 与 p_{i-2} 之差数的一半. 易知差数在符号方面是一正一反的. 令 d_i 表示差数 p_i-p_{i-1}，则对一切 i，均有 $d_i=-\frac{1}{2}d_{i-1}$. 这些事实，均表

示于图 2 中.

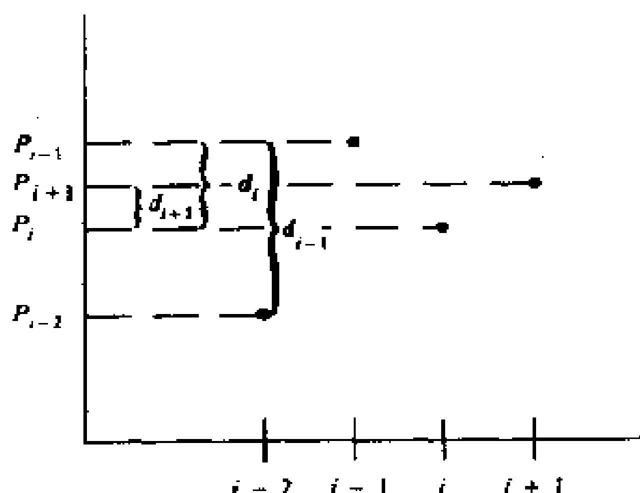


图 2

各个相继 p 值之间的关系.

反复应用关系式 $d_i=-\frac{1}{2}d_{i-1}$，我们将得出 $d_i=-\frac{1}{2}d_{i-1}=-\frac{1}{2}\left(-\frac{1}{2}d_{i-2}\right)=\cdots=\left(-\frac{1}{2}\right)^{i-2}d_2$. 但 d_2 等于 p_2-p_1，即$\frac{3}{4}-\frac{1}{2}=\frac{1}{4}$. 因而 d_i 等于$\frac{1}{4}\left(-\frac{1}{2}\right)^{i-2}$，即$\left(-\frac{1}{2}\right)^{i}$. 最后，利用下列事实，即 p_i 可以记作

$$(p_i-p_{i-1})+(p_{i-1}-p_{i-2})+\cdots-(p_2-p_1)+p_1$$

或 $d_i+d_{i-1}+\cdots+d_2+p_1$，通过直接代入，可以算得 p_i 等于$\left(-\frac{1}{2}\right)^{i}+\left(-\frac{1}{2}\right)^{i-1}+\cdots+\left(-\frac{1}{2}\right)^{2}+\frac{1}{2}$. 利用熟知的几何级数有限项求和公式即可求出其结果

$p_i=\frac{2}{3}+\frac{1}{3}\left(-\frac{1}{2}\right)^{i}$. 在表 1 中给出了前面 12 个 p 值.

此时，可以观察到一桩有趣事实，即 p 值序列有规律地环绕着$\frac{2}{3}$这个数进行振荡，当人们沿着该数列前进时，振幅将会越来越小. 其次，有奇数下标的子序列 $p_1,p_3,p_5,\cdots$ 单调递增地趋向极限值$\frac{2}{3}$，而其

互补子序列 $p_2, p_4, p_6, \cdots$ 则单调递减地趋向同一极限值. 由此可见，最大的 p 值是 p_2，次大的是 p_4，等等，以此类推；而最小的 p 值是 p_1，次小的是 p_3 等等.

i	p_i
1	0.5000
2	0.7500
3	0.6250
4	0.6875
5	0.6563
6	0.6719
7	0.6641
8	0.6680
9	0.6660
10	0.6670
11	0.6665
12	0.6667

表 1

本游戏的简化模型中前十二个概率值 p_i.

现在我们已经可以答复这位教师的问题（至少对本游戏的简化模型来说）. 为了使期望的作业数最多，应当把作业布置在偶数的格子里，从第 2 格开始，一直序贯地布置下去，直到发生以下两种情况之一时为止. (a)所有的作业都已布置完毕. (b)已经到了棋盘的终点. 在(b)这种情形，余下来的作业应放在未占据的空格（即奇数番号的格子）中，从末尾开始，按相反的顺序进行.

安排得体，获益多少？

作为一名带有应用偏好的数学工作者，我想知道把作业指令最

佳地布置在棋盘上到底有多少好处．把问题说得更确切一些，就是要问，在最好的布置作业法与最坏的布置作业法（只要在最佳方案中把奇、偶数的任务颠倒一下就行）之间，作业期望数之差究竟是多少？使用一些极为简单的代数，即可给出问题的答案．其结果取决于 n（棋盘上的格子数）是否至少为 t（作业的个数）的二倍，如果 $n \geqslant 2t$，则答案是 $(1-4^{-t})/3$ 或 $(1-4^{t-n})/3$．不论 n 或 t 有多大，这个数决不能超过 $\frac{1}{3}$，在绝大多数情况下，此数非常接近于 $\frac{1}{3}$．譬如说，如果一共有 5 个作业，棋盘上有 10 格或更多格子，则在最优布置法中，作业的期望数为 3.4443，而在最差布置法中，期望数是 3.1113，两者差数为 0.3330（确切地说是 341/1024）．虽然这个差数看来似乎无足轻重，但若反复进行这种游戏，它就会显得越来越重要．例如，假定一个 30 人班级的每位学生都来做一次游戏，那么最优布置法与最差布置法之间的差别就大致相当于 10 个外加的阅读作业能否执行．

原先的问题

在解决了较简单的问题之后，让我们尝试一下，把同样的办法应用到这位教师原来提出的问题，也就是，允许走一格、二格或三格的情况．通过同样的论证法可以得出一个递推关系式．此时，关系式变为 $p_i=(p_{i-1}+p_{i-2}+p_{i-3})/3$（当 i 大于 3 时）．为了把计算贯彻下去，我们需要算出 p_1，p_2，与 p_3．显然，p_1 等于 $\frac{1}{3}$，走到第 2 格只有两种可能性，要末是第一步走两格，要末第一、二步都各走一格，因此 p_2 等于 $\frac{1}{3}+\left(\frac{1}{3}\right)^2$，即 $\frac{4}{9}$．走到第 3 格有四种方式：(1)一步走 3 格；(2)一步走 2 格，下一步再走 1 格；(3)一步走 1 格，下一步再走 2 格；(4)三步都各走 1 格．由此可见 p_3 等于 $\frac{1}{3}+\left(\frac{1}{3}\right)^2+\left(\frac{1}{3}\right)^3=\frac{16}{27}$．递推关系式，再加上这三个初始值，就能使我们得以在原则上算出任何一个 p_i．在表 2 中给出了前 12 个 p 值．

i	p_i
1	0.3333
2	0.4444
3	0.5926
4	0.4568
5	0.4979
6	0.5158
7	0.4902
8	0.5013
9	0.5024
10	0.4980
11	0.5006
12	0.5003

表 2

女教师原先的游戏. 前 12 个概率值 p_i.

但是，不像上述简单游戏，要想从递推关系式推出直接计算 p_i 的公式是没有简单方法可以遵循的[1]. 尽管如此，还是可以得出一些有趣的观察结果. 由于递推关系式表明，第三个之后的任一 p 值都是前面三个值的平均数；所以它必然要大于三数中的最小一数而小于最大一数. 反复应用这一论证的结果，人们可以断言，在序列 $\{p_1, p_2, \cdots\}$ 中的任意三个相继值，其最大的一数必定大于序列中后面的一切项，而三数中的最小者则应小于序列中其后的一切项. 这一注解的直接后果是：p_1, p_2, p_3 中的最大数 p_3 是一切 p 值中的最大者，而这三数中的最小数 p_1，则是所有 p 值中的最小者.

p 值的分布是否像简化模型那样，表现出某种规律性？也许你会

---

[1] 原注：具有较高水平的读者也许愿意了解 p_i 可表示为下式：$\frac{1}{2}+\frac{(-1)^i}{4}\left[\left(1+\sqrt{-2}\right)^i+\left(1-\sqrt{-2}\right)^i\right]$.

情不自禁地去推测，最大的 p 值将按递降顺序排列，即由子序列 p_3，p_6，p_9，…来给出. 可惜实际情况根本不是那样，察看一下表 2 就会知道，所谓的规律在 p_{12} 的地方就遭到破坏，p_{12}（=0.5003）并不是第四个最大的值，而只不过是名列第六，它小于 p_8（=0.5013）与 p_{11}（=0.5006）. 与此相似，把较小的 p 值按上升顺序排列时也看不出什么规律性，它们是：p_1，p_2，p_4，p_7，p_5，p_{10}，等等.

表 2 确实在提示人们，p 值的序列收敛于极限值$\frac{1}{2}$；事实上，这是可以证明的，但其证法需要相当高深的数学知识. 作为这一事实的简单推论，不可能有接连三个 p 值位于$\frac{1}{2}$的同一侧. 对此，我们可以使用反证法，假定真的有那样三个 p 值存在，则根据上面的注解，在其后的一切 p 值都必然介于三数中的最大者与最小者之间. 因此它们必然要比三数中最接近$\frac{1}{2}$的那个数更为远离$\frac{1}{2}$. 这就与 p 值收敛于$\frac{1}{2}$的话产生了矛盾.

p 值永远在数$\frac{1}{2}$的上下振荡，但决不会有两个以上的 p 值位于$\frac{1}{2}$的同侧. 可是，对于振荡来说，看不出什么明显的规律性. 表 3 给出了前面 25 个 p 值的振荡情况.

i	1	2	3	4	5	6	7	8	9	10	11	12	
p_i 对比 0.5	−	−	+	−	−	+	−	+	+	−	+	+	
i	13	14	15	16	17	18	19	20	21	22	23	24	25
p_i 对比 0.5	−	+	+	−	+	−	−	+	−	−	+	−	−

表　3

前 25 个 p 值在 0.5 的上下位置. + 表示在上，− 表示在下.

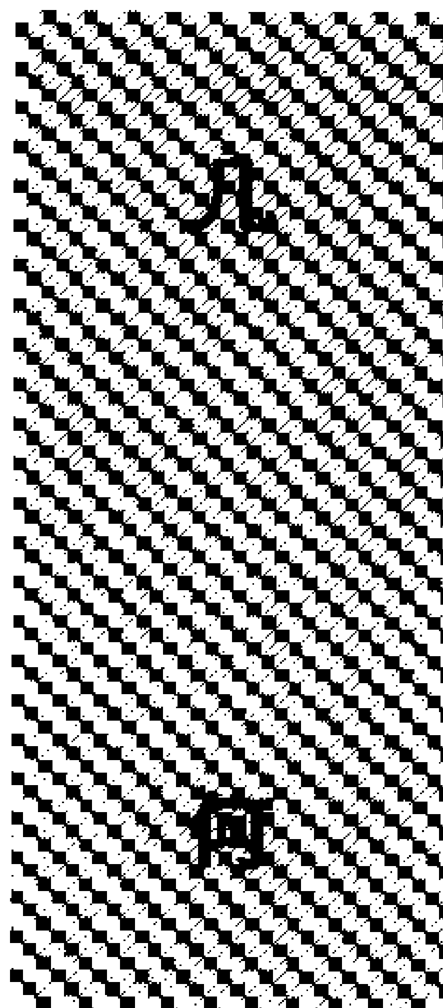
现在设想我们要把已学到的知识用于以上的特定例题. 如果有五个作业要安排在棋盘上，而棋盘上至少有十一格，则最好的安排法是第 3，6，8，9，11 格，这将给出做作业的期望值 2.6127，与此相反，

最坏的安排是放在第 1,2,4,5,7 格,给出的相应期望值为 2.2226. 好、坏之差是 0.3901,有趣的是,这一差别要比简化模型的好、坏差别为大.

最后几句话

这个问题表明,一种解法的某些方面可以推广到一切场合,而某些方面却只局限于所考察的特定问题. 对 p_i 导出递推关系的论证对本游戏的各种变化形式都是有效的,但解法的具体模式则要强烈依赖于可能的走法究竟有多少种. 当然,两种最常见的随机独脚跳游戏(可允许走 1 格至 6 格或 2 格至 12 格)在这里并没有触及,但读者们现在也许已处在能藉助计算机加以分析的有利地位.

读通这篇文章之后,再走他自己的路,读者们也许能得到推广其结论的乐趣. 说到底,只知工作而不懂玩耍,再聪明的孩子也会变笨!


凡例

相切圆的花环

● 南加利福尼亚大学

所罗门·W·果隆姆(Solomon W. Golomb)

众所周知,一个圆恰能被六个与它同样大小的圆环绕(见图 1).一般地说,半径为 s 的 n 个等圆可以密合无间地环绕一个半径为 r 的圆.这里,$n\geqslant 3$,根据初等三角学,我们有如下关系式(见图 2)

$$\sin\frac{\pi}{n}=\frac{s}{r+s},$$

由此 $r=s\left(\csc\frac{\pi}{n}-1\right)$.

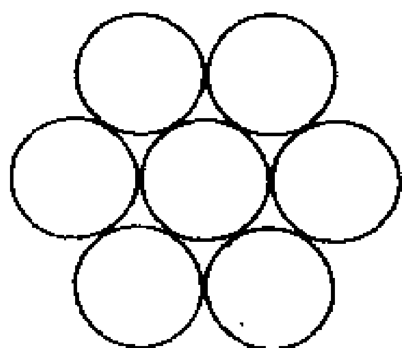


图 1

一个圆的周围环绕着六个等圆.

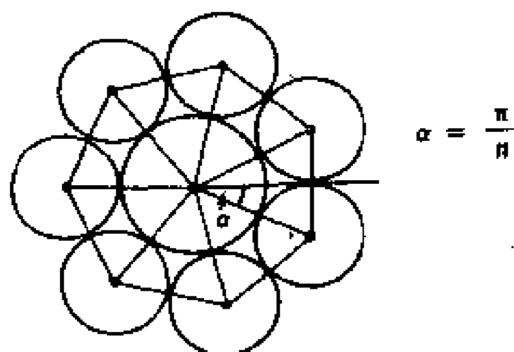


图 2

环绕一个给定圆的 n 个等圆,$n=7$的情形.

如将半径固定为 $r=1$，则当 n 递增时，s 递减，如表 1 所示.

n	s
3	$6.46410=3+2\sqrt{3}$
4	$2.41421=1+\sqrt{2}$
5	1.42592
6	1.00000
7	0.76642
8	0.61991
9	0.51980
10	0.44721

表 1

密合无间地环绕单位圆的、半径为 s 的 n 个等圆.

我们下面将研究内圆与环绕它的 n 个圆的半径可以各不相等的情况(见图 3). 当人们手中有一系列大小不等的圆形硬币，企图用若干个硬币环绕一个硬币时，就会遇到这个问题. 首先，我们要对 n 个圆密合无间地环绕一个给定圆的提法给出一个确切的定义.

定义 圆 C^* 称为被 n 个圆 $C_1,C_2,\cdots,C_n$ 密切环绕，如果每个圆 C_i 都外切于 C^*，且又与 C_{i-1} 及 C_{i+1} 分别外切的话(各下标作为模 n 的同余看待).

设周围各圆的半径为 $r_1,r_2,\cdots,r_n$，则被包围的圆 C^* 的半径有无一个简单公式去计算?

令人惊讶的是，$n>3$ 时，被包围之圆的大小不仅取决于周围环绕诸圆之半径，而且与各圆的相邻顺序有关. 一般来说，给定 n 个外围圆，则被围内圆的大小，可多达 $\frac{1}{2}(n-1)!$ 种，取决于外围诸圆的排列顺序. (事实上，如果外围诸圆的半径是 n 个代数独立的实数，则被包围的内圆半径，确可存在 $\frac{1}{2}(n-1)!$ 个不同的值.)外围诸圆在先后顺序上的重要性，我们在图 4 中作了例示.

在图 4 中，图 C_6 与 C_6' 的大小相同，可是 C_6 放置在 C_4 与 C_5 之间，于是在包围 C^* 时起了作用. 然而 C_6' 却被放置于 C_1 与 C_2 之间，

对包围 C^* 的任务无所贡献. 如果把 C_6' 稍为放大一点,它将迫使 C_1 与 C_2 分离,然而,尽管如此,它所起的作用仍然没有当与它相邻的两圆大小和它比较接近时来得大!

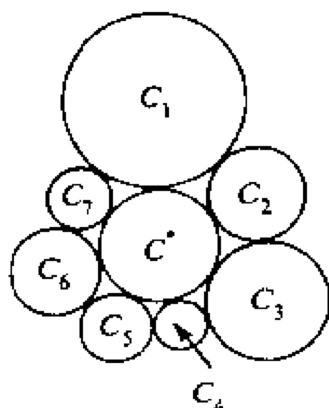


图 3

一个圆被七个不相等的圆密切包围.

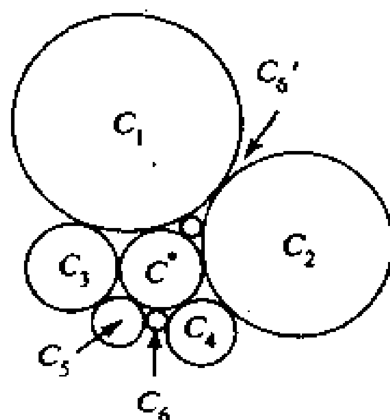


图 4

圆 C_6 与 C_6' 大小相等,但在包围 C^* 时所起的作用极不一样.

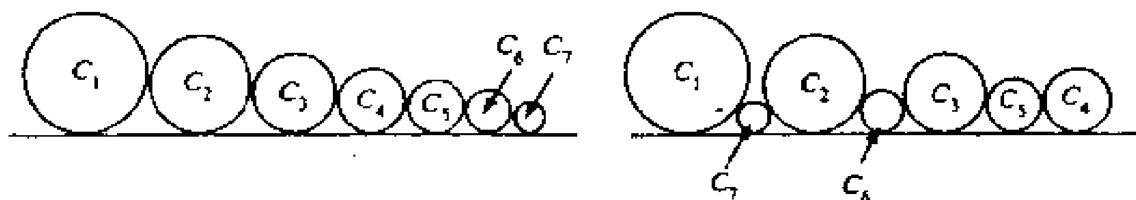


图 5

将 n 个不等圆放置于直线上的极大化与极小化策略.

粗略地说,如果外围诸圆的大小尽量接近,则被围的内圆半径将有可能是极大值. 反之,如果外围诸圆的大小相差悬殊,则内圆半径将有可能是极小. 如果我们的任务是只要求把 n 个圆沿直线排成一串,则极大化与极小化策略就表现得明显了(图 5). (直线可视为一个具有无限长半径的内圆,它当然不可能被个数有限、大小有限的圆所紧密包围.) 我们可以把诸圆编号如下:$C_1, C_2, C_3, \cdots, C_n$,使其相应半径满足关系式:$r_1 \geqslant r_2 \geqslant r_3 \geqslant \cdots \geqslant r_n$.

注: 如果 C_n 小于与圆 C_1, C_2, C^* 同时相切的圆,则“极小化”的

说法是尚未明确定义的. 不过,我们眼下将不考虑这一极端情形的极小化问题.

至于怎样放置 n 个外圆,使被围绕的内圆半径最大的问题,图 6 提供了一个经验算法. 从 C_1 起,一个方向是 $C_2, C_4, C_6, \cdots\cdots$,另一方向则为 $C_3, C_5, C_7, \cdots\cdots$. 使内圆 C^* 的半径最小的一个经验算法则如图 7 所示,从 C_1 起,一个方向是 $C_n, C_2, C_{n-2}, C_4, C_{n-4}, C_6, \cdots\cdots$,而另一方向则是 $C_{n-1}, C_3, C_{n-3}, C_5, C_{n-5}, \cdots\cdots$. 之所以称之为经验算法,是因为对诸圆 $C_1, C_2, C_3, \cdots\cdots, C_n$ 半径的各种可能选择,这两个算法的极值性质还没有被满意地证明或推翻过.

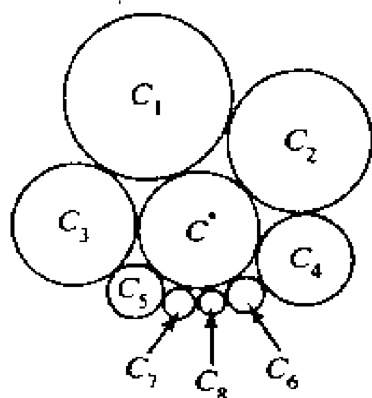


图 6

使 C^* 的半径最大的经验算法.

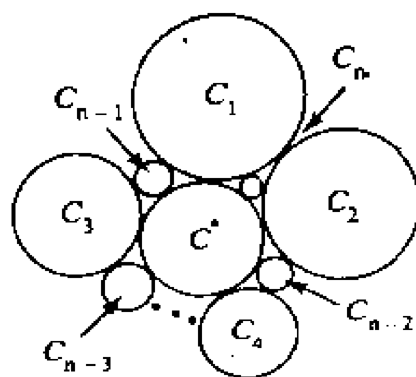


图 7

使 C^* 半径最小的经验算法.

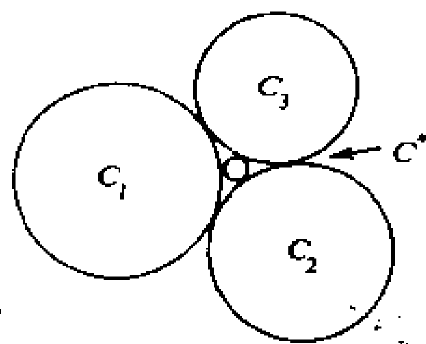


图 8

外围三圆唯一地确定了 C^*.

在 $n=3$ 的情形,除了欧氏刚体运动(旋转,反射,平移)之外,仅有一种办法来配置三个圆 C_1, C_2, C_3 使之互相外切,而这样也就唯一地确定了被包围的圆 C^* 的半径(见图 8). 设 C_1, C_2, C_3 的半径分别是 a, b, c,则圆 C^* 的半径可由下式表出:

$$r = \frac{abc}{ab + bc + ca + 2\sqrt{abc(a + b + c)}}.$$

$n = 4$ 时，如果圆 C_1, C_2, C_3, C_4 的半径均不相等，则有 $\frac{1}{2}(4-1)! = 3$ 种实质不同的方式来包围内圆 C^*，如图 9 所示.

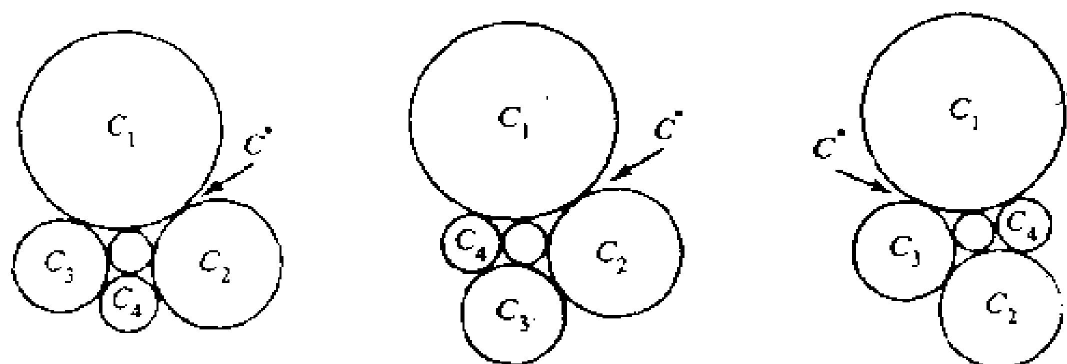
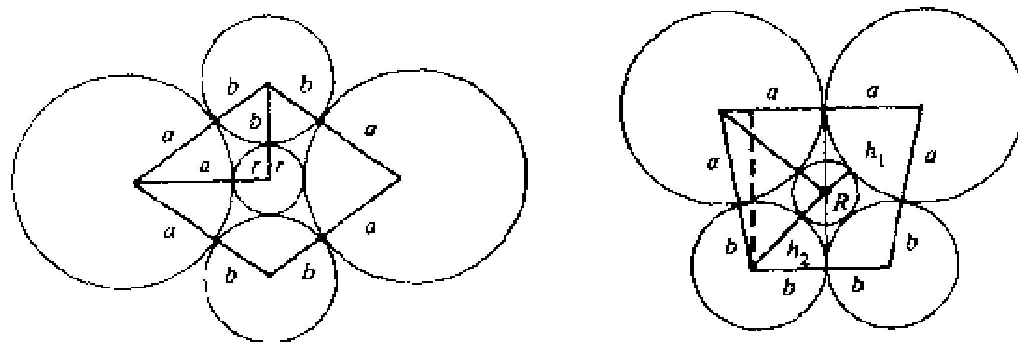


图　9

$n = 4$ 时，三种实质不同的配置顺序.

我们来讨论外围四圆中只有两种相异半径的特殊情形. 设圆 C_1, C_2 的半径为 a；C_3, C_4 的半径为 b，此时，仅有的两种不同配置方式，见图 10 所示.



第 I 种情形　　　　第 II 种情形

图　10

由半径为 a 的两圆与半径为 b 的两圆包围一个内圆的两种不同方式.

在第 I 种情形，由毕达哥拉斯定理，我们有 $(a+b)^2=(a+r)^2+(b+r)^2$，由此得 $2ab=2ar+2br+2r^2$，即 $r^2+(a+b)r-ab=0$，于是由二次方程求根公式，得

$$r=\frac{\sqrt{a^2+6ab+b^2}-(a+b)}{2}.$$

对 a,b 的某些整数值可求得 r 的整数值. 例如 $r(3,2)=1$，$r(10,3)=2$，$r(12,5)=3$ 等等.（它们与毕达哥拉斯三数组的关系如下：若 $r(a,b)=r$，则 $(b+r,a+r,a+b)$ 是毕达哥拉斯三数组，这可由图 10 的第 I 种情形得到说明. 反之，如 (A,B,C) 是一个毕达哥拉斯三数组，即 $A^2+B^2=C^2$ 得到满足，则我们可得出情形 I 的不定方程的解：$a=(A-B+C)/2$，$b=(-A+B+C)/2$，$r=(A+B-C)/2$.）有趣的是，半径为 1 的一个硬币得以被两个半径为 2 的硬币与两个半径为 3 的硬币紧密围住，只要把它们按照图 10 的第 I 种情形配置；另一方面，如果它们按第 II 种情形配置，那么虽然看上去也像密合无间，但我们将可以证明，实际情况并非如此.

对第 II 种情形，令 $h=h_1+h_2$，由毕达哥拉斯定理得

$$h_1^2=(a+R)^2-a^2=2aR+R^2,$$

$$h_2^2=(b+R)^2-b^2=2bR+R^2,$$

另外我们注意到图上虚线的长度为 h，则有 $h^2=(a+b)^2-(a-b)^2=4ab$，于是

$$4ab=h^2=(h_1+h_2)^2=h_1^2+h_2^2+2h_1h_2=2(aR+bR+R^2+h_1h_2),$$

由此，

$$h_1h_2=2ab-(aR+bR+R^2),$$

$$(2aR+R^2)(2bR+R^2)=h_1^2h_2^2=[2ab-(aR+bR-R^2)]^2,$$

$$(2a+R)(2b+R)=\left(\frac{2ab}{R}-(a+b)-R\right)^2,$$

$$4ab+2(a+b)R+R^2=\frac{4a^2b^2}{R^2}+(a+b)^2+R^2-4ab+2(a+b)R$$

$$-\frac{4ab(a+b)}{R},$$

$$R^2(a^2-6ab+b^2)-4ab(a+b)R+4a^2b^2=0.$$

由二次方程求根公式，可得

$$R=\frac{4ab(a+b)\pm\sqrt{16a^2b^2\{(a+b)^2-(a^2-6ab+b^2)\}}}{2(a^2-6ab+b^2)},$$

因为 R 必须大于 0，于是有

$$R=\frac{2ab\{2\sqrt{2ab}-(a+b)\}}{8ab-(a+b)^2}=\frac{2ab}{(a+b)+2\sqrt{2ab}}.$$

由此我们看到，对于取整数值的 a 与 b，当且仅当 $2ab$ 是完全平方数时，R 才是有理数. 特别地，$R(3,2)=\frac{12}{5+4\sqrt{3}}=\frac{12}{23}(4\sqrt{3}-5)=1.006019^-$，也就是说，比 $r(3,2)=1$ 增加了大约 0.6%. 这个差别是如此微小，以致在实践中几乎不为人注意. 然而，当比值 a/b 递增时，比值 R/r 也递增. 例如，$r(10,3)=2$，而 $R(10,3)=\frac{60}{13+4\sqrt{15}}=\frac{60}{71}(4\sqrt{15}-13)=2.10586$，也就是说，内圆 C^* 的半径 R 比 r 差不多要增大 5%.

当圆 C^* 被一个半径为 b 的圆与 n 个半径为 a 的圆紧密包围时，外围诸圆的排列顺序不会影响圆 C^* 的半径. 当 $n=2$ 时，这是一个四圆相互(外)切的特例. 其一般情形是，半径为 r 的圆 C^* 被半径依次为 a,b,c 的三个圆所紧密包围. 这在前面已提到过，而在文献〔1〕中有详细推导，这种一般情形下的公式为

$$r=\frac{abc}{ab+bc+ca+2\sqrt{abc(a+b+c)}}=\frac{S_3}{S_2+2\sqrt{S_1S_3}},$$

这里 $S_1=a+b+c$，$S_2=ab+bc+ca$，$S_3=abc$ 是 a,b,c 的三个初等对称函数.

在我们这个特例中，$a=c$，当然是更加简单，并能满足

$$a=\frac{4br(b+r)}{(b-r)^2},\quad b>r.$$

图 11

被两个半径为 a 的圆与一个半径为 b 的圆紧密包围的圆.

图 11 告诉我们：

$$(a+r)^2 = a^2 + x^2, (a+b)^2 = a^2 + (x+r+b)^2.$$

于是，

$$x + r + b = \sqrt{2ab + b^2},$$

其中

$$x = \sqrt{2ar + r^2},$$

故有

$$r + b = \sqrt{2ab + b^2} - \sqrt{2ar + r^2},$$

$$r^2 + 2rb + b^2 = (2ab + b^2) + (2ar + r^2) - 2\sqrt{(2ab + b^2)(2ar + r^2)},$$

$$(ab + ar - rb)^2 = (2ab + b^2)(2ar + r^2),$$

$$a^2b^2 + a^2r^2 = 2a^2br + 4abr^2 + 4arb^2,$$

$$(b - r)^2 = (1/a)(4br)(b + r),$$

$$a = 4br(b + r)/(b - r)^2.$$

b 与 r 的明显对称性在几何上并未能予以实现，这是由于显然 $b > r$ 之故. 如果 a 与 b 的数值已给出，则由二次方程求根公式可得

$$r = b\frac{(a + 2b) - 2\sqrt{b^2 + 2ab}}{a - 4b}.$$

$a=4b$ 的情形是一个"可移去的奇点"，因为在此处 $r=\frac{1}{3}b$. 然而，在有关的方程

$$b=r\frac{(a+2r)+2\sqrt{r^2+2ar}}{a-4r}$$

中，$a=4r$ 却是一个真正的奇点，对应于 $b=\infty$. 如果设 $r=1,a=4$，我们从图 12 中的 3—4—5 直角三角形看出，三个两两外切的圆同时也与一条直线相切（直线可视为 $b=\infty$ 的圆）.

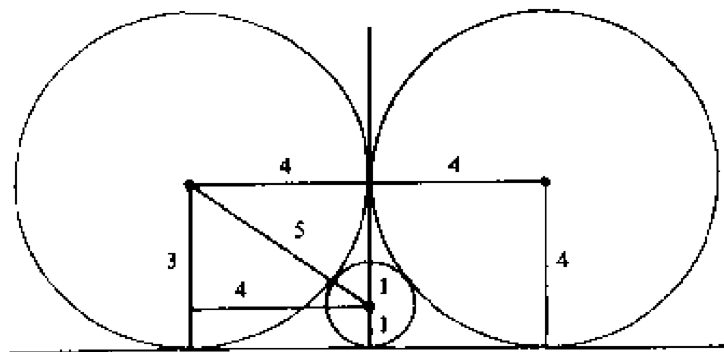


图 12

半径为 1 的圆被一条直线与两个半径为 4 的圆紧密包围.

有许多取整数值的三数组 (r,b,a)，其中包括 $(1,2,24)$，$(1,3,12)$，$(5,7,420)$，$(6,14,105)$ 等等，它们可从以下一些式子得出：

1. $r=n,b=n+1,a=4n(n+1)(2n+1)$；
2. $r=n,b=n+2,a=2n(n+1)(n+2)$；
3. $r=2n,b=2n+8,a=n(n+2)(n+4)$ 等等.

让我们也来讨论被一个半径为 b 的圆与 n 个半径为 a 的圆紧密包围且其半径 r 也是 b 的圆 C'. 对任一 $n\geqslant 3$，这一种构形唯一地决定了一个比值 $Q_n=a/b$. 我们将稍为详细地对 $n=3,4,5,6$ 的情形研讨这个 Q_n 的数值.

在图 13 中，我们看到的是 $n=3$ 的情形. 我们有

$$(2a)^2=x^2+(a+2b)^2 \text{ 与 } (a+b)^2=x^2+b^2.$$

于是，$(2a)^2-(a+b)^2=(a+2b)^2-b^2$，由此而得 $a^2-3ab-2b^2=$

0．利用二次方程求根公式，可算出 $Q_3 = a/b = (3+\sqrt{17})/2 = 3.5615528\cdots$.

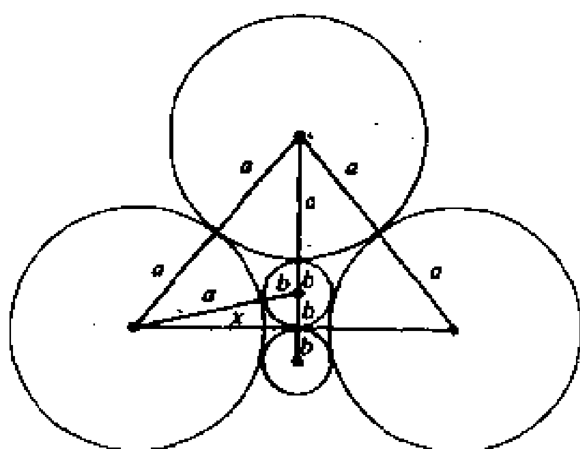


图 13

一个半径为 b 的圆被三个半径为 a 的圆与一个半径为 b 的圆紧密包围.

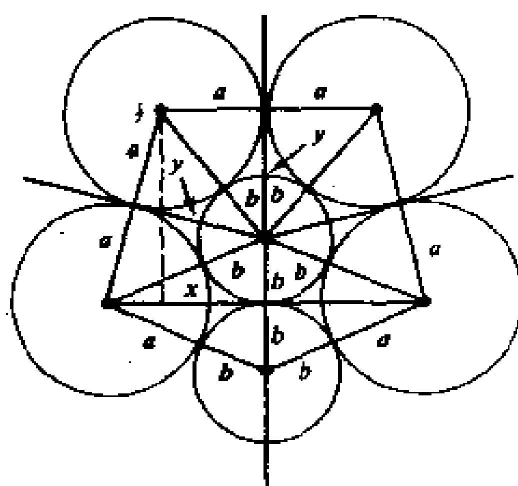


图 14

半径为 b 的一个圆被四个半径为 a 的圆与一个半径为 b 的圆紧密包围.

$n=5$ 时，显然 $Q_5 = a/b = 1$，这是因为它归结为图 1 的那种构形. 可是，当 $n=4$ 时，情形就要稍为复杂一些，从图 14 中，我们可以

观察到：

$$(a+b)^2=b^2+x^2,$$
$$(a+b)^2=a^2+y^2,$$
$$(2a)^2=(y+b)^2+(x-a)^2.$$

于是，$x=\sqrt{a^2+2ab}$，$y=\sqrt{b^2+2ab}$，$a^2-2ab-b_2=b\sqrt{b^2+2ab}-a\sqrt{a^2+2ab}$，由此，得

$$3a^2-ab-b^2=\sqrt{(a^2+2ab)(b^2+2ab)},$$

经整理得 $9a^4-8a^3b-10a^2b^2+b^4=0$．所以 $Q_4=a/b$ 是方程 $9x^4-8x^3-10x^2+1=0$ 的一个根，其数值解是 $Q_4=1.5684897\cdots$（所有四次方程的根原则上都可以用根式表示，但对本例来说，用根式表为显式是过于复杂了，故予省略）.

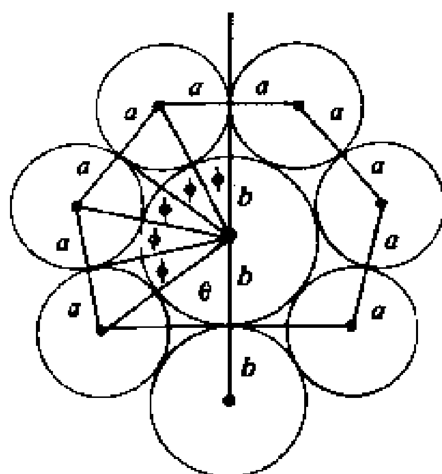


图 15

当半径为 b 的一个圆被 n 个半径为 a、一个半径为 b 的圆紧密包围时，显然有

$$\pi=\theta+(n-1)\varphi\text{，此处 }\theta=\cos^{-1}\frac{b}{a+b},\varphi=\sin^{-1}\frac{a}{a+b}.$$

$n=6$（见图 15）的情形是在加莱·A·福特(Gary A. Ford)于1973年致马丁·加德纳的一封未公开的信件中首先作为问题而提出的. Ford 补充说，该问题是由排列硬币而引起(例如一角与二角五分两种币)，他同他在马里兰大学的同事们所能做到的最好结果只是

把 Q_6 表为一个十次多项式的根(福特后来把这个问题发表于麻省理工学院的《Technology Review》上,参看文献[2]).利用上面已经说明过的方法,我们是否也能办到或者做得更好些?更一般地说,对所有的 $n\geqslant 3$,是否有可能把 Q_n 表示为 n 的代数或三角函数呢?我们将会看到,一般三角表达式确实存在,由此可导出一个代数方程.作为其特例,我们将把 Q_6 表为一个 8 次多项式的根.

在图 15 中,我们可以看到,以 b 为半径的内圆中的一个平角是角 θ 与角 φ 的五倍之和,此处 $\theta=\cos^{-1}\frac{b}{a+b}$ 而 $\varphi=\sin^{-1}\frac{a}{a+b}$.对一般的 n,此结果将是

$$\pi=\cos^{-1}\frac{b}{a+b}+(n-1)\sin^{-1}\frac{a}{a+b}.$$

如设 $a/(a+b)=\alpha$,$b/(a+b)=\beta$,则 $\alpha+\beta=1$,而比值 $Q_n=a/b=\alpha/\beta$,且有

1. $\pi=\cos^{-1}(1-\alpha)+(n-1)\sin^{-1}\alpha$.

从实际的计算方面说,此公式已足以把 α(由它可算出 Q_n)算到任意精度.不过,对任一 n,也有可能把方程 1 代之以一个以 α 为根的代数方程.此外,也有可能导出以 $Q_n=\frac{\alpha}{1-\alpha}$ 为其根的代数方程,因为若 $f(x)=0$ 有一根为 $x=\alpha$,则通过直接代入,易于证明 $g(x)=f\left(\frac{x}{1+x}\right)=0$ 有一根为 $\frac{\alpha}{1-\alpha}$.

办法是把 1 式重新改写为 $(n-1)\sin^{-1}\alpha=\pi-\cos^{-1}(1-\alpha)$,然后在等式的两边各取余弦,即

$$\cos[(n-1)\sin^{-1}\alpha]=\cos[\pi-\cos^{-1}(1-\alpha)]=\alpha-1.$$

令 $\sin^{-1}\alpha=z$,则大家知道,$\cos(n-1)z$ 是 $\cos z$ 的一个 $n-1$ 次多项式,且 $\cos z=\cos(\sin^{-1}\alpha)=\sqrt{1-\alpha^2}$.于是,在最不利的情况,$\alpha$ 可能是一个 $2(n-1)$次多项式的根.实际上,若 n 为奇数,对所有的 $n\geqslant 3$,Q_n 满足一个次数$\leqslant n-1$ 次的代数方程.若 n 为偶数,则对一切 $n\geqslant 4$,Q_n 满足一个次数$\leqslant 2(n-2)$的代数方程.

我们将对(已解决的)$n=3$ 与 $n=4$ 进行实际计算以资说明.

当 $n=3$ 时，$\cos 2z=\alpha-1$，$2\cos^2 z-1=\alpha-1$，$2\cos^2 z=\alpha$，$2(1-\alpha^2)=\alpha$，$2\alpha^2+\alpha-2=0$，即 α 满足 $f(x)=2x^2+x-2=0$，于是 Q_3 应满足 $g(x)=f\left(\dfrac{x}{1+x}\right)=0$，$(1+x)^2 g(x)=2x^2+x(1+x)-2(1+x)^2=x^2-3x-2=0$，故 Q_3 是方程 $g(x)=x^2-3x-2=0$ 的一个根，其值为 $\dfrac{3+\sqrt{17}}{2}=3.56155\cdots$.

类似地，当 $n=4$ 时，$\cos 3z=\alpha-1$，$4\cos^3 z-3\cos z=\alpha-1$，$\sqrt{1-\alpha^2}[4(1-\alpha^2)-3]=\alpha-1$，$(1-\alpha^2)(1-4\alpha^2)^2=(\alpha-1)^2$，$(1+\alpha)(16\alpha^4-8\alpha^2+1)=1-\alpha$，$16\alpha^5+16\alpha^4-8\alpha^3-8\alpha^2+2\alpha=0$，由于 $\alpha=0$ 显然不是一个可能的解，所以 α 是 $f(x)=8x^4+8x^3-4x^2-4x+1=0$ 的一个根，因而 $Q_4=\dfrac{\alpha}{1-\alpha}$ 应满足方程 $g(x)=f\left(\dfrac{x}{1+x}\right)=0$，故有

$$(1+x)^4 g(x)=8x^4+8x^3(1+x)-4x^2(1+x)^2-4x(1+x)^3+(1+x)^4=9x^4-8x^3-10x^2+1=0,$$

从而 Q_4 的数值解为 $Q_4=1.56849\cdots$.

在附表 2 中，我们给出了 $3\leqslant n\leqslant 9$ 时 Q_n 所满足的多项式，并给出了相应的 Q_n 值. 从中已能看出这些多项式的系数模式. n 的奇、偶明显地分别对应于大相径庭的多项式集群.

最后我们要提一下，当三圆两两外切并都切于一直线时，记三圆的半径为 a,b,c 而且 $a\geqslant b\geqslant c$，则有一个美妙关系式：

$$\frac{1}{\sqrt{a}}+\frac{1}{\sqrt{b}}=\frac{1}{\sqrt{c}},$$

(这推广了图 12 中的情况)请参阅文献[3].

n	多项式次数	Q_n 应满足的多项式	Q_n 的值
3	2	x^2-3x-2.	3.56155
4	4	$9x^4-8x^3-10x^2+1$.	1.56849
5	4	$x^4-11x^3+x^2-7x+2$.	1.00000
6	8	$25x^8-188x^7+236x^6+436x^5-2x^4-180x^3$	0.73403

（续表）

n	多项式次数	Q_n 应满足的多项式	Q_n 的值
		$-68x^2-4x+1.$	
7	6	$x^6-31x^5+40x^4+42x^3-7x^2-11x-2.$	0.58027
8	12	$49x^{12}-956x^{11}+5090x^{10}-3036x^9-11121x^8$	0.48015
		$+1800x^7+10140x^6+4200x^5-865x^4+972x^3$	
		$-222x^2-12x+1.$	
9	8	$x^8-55x^7+259x^6+77x^5-215x^4-101x^3$	0.40977
		$-17x^2+15x+2.$	

表 2

Q_n 的多项式与其值，$3\leqslant n\leqslant 9$.

表 2 的附注

1. 对具有 $4k+1$ 形状的 n，我们所得出的 $n-1$ 次多项式似乎总有一根 $x=1$. 对 $n=5$，这对应于 $Q_5=1$. 然而，对 $n=9,13,17,\cdots$，Q_n 所满足的方程，其次数却是 $\leqslant n-2$.

2. 对 $n>4$，这些多项式没有一个已被证明为既约多项式. 请注意，$n=5$ 时，该多项式是 $x-1$ 与既约三次多项式 x^3-10x^2-9x-2 之积. $n=9$ 时，分解式是：

$$(x-1)(x^7-54x^6+205x^5+282x^4+67x^3-34x^2-17x-2).$$

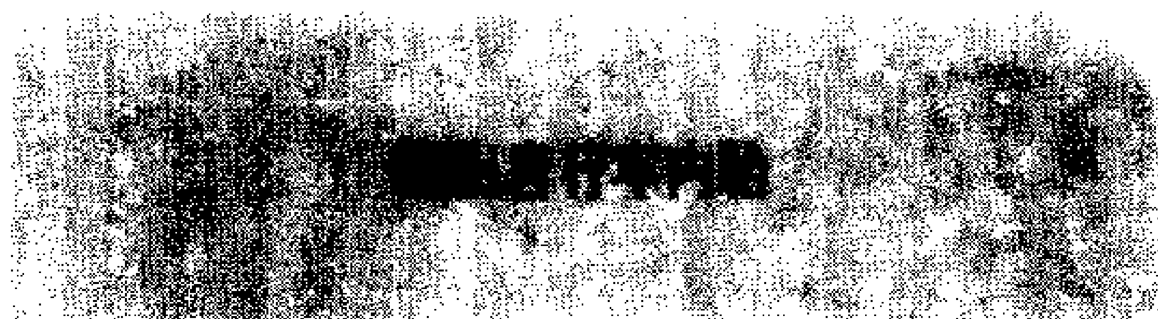
参 考 文 献

1 Beecroft, Philip. 1842. Properties of circles in mutual contact. *Lady's and Gentleman's Diary*, pp. 91-96.

2 Ford, Gary A. 1974. *Technology Review*, Problem June 5, vol. 76: 57-8. (See also problem NS 13, vol. 81, November 1978, p. 84.)

3 Trigg, C. W. 1940. Problem E432, *American Math. Monthly*, 47: 487.

4 _____. 1941. Solution to Problem E432, *American Math. Monthly*, 48: 267-68.



● 南加利福尼亚大学

[美] 赫伯特·泰勒(Herbert Taylor)

古老的橡皮体几何学研究那些只要保持光滑、完整，可以任意变形、弯曲、延伸与扭转的曲面. 一个很有名的拓扑游戏是要人们想象出脚踏车的内胎翻出来后将是什么样子. 据我所知，这类妙技在严肃的意义上同数学无甚瓜葛，但它们可用来培养灵活的形象思维能力.

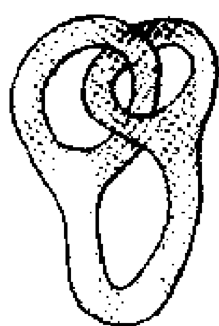


图 1A

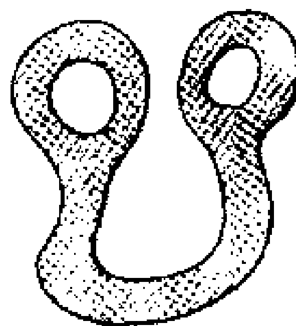


图 1B

作为第一个例子，让我们来看一下，图 1A 的曲面怎样变为图 1B. 要求读者进行想象，或绘出一系列的图形，从图 1A 变为图 1B，在变形过程中不准切割曲面，也不能让曲面的一部分接触另一部分. 图 1C 给出了一种可能的变换系列.

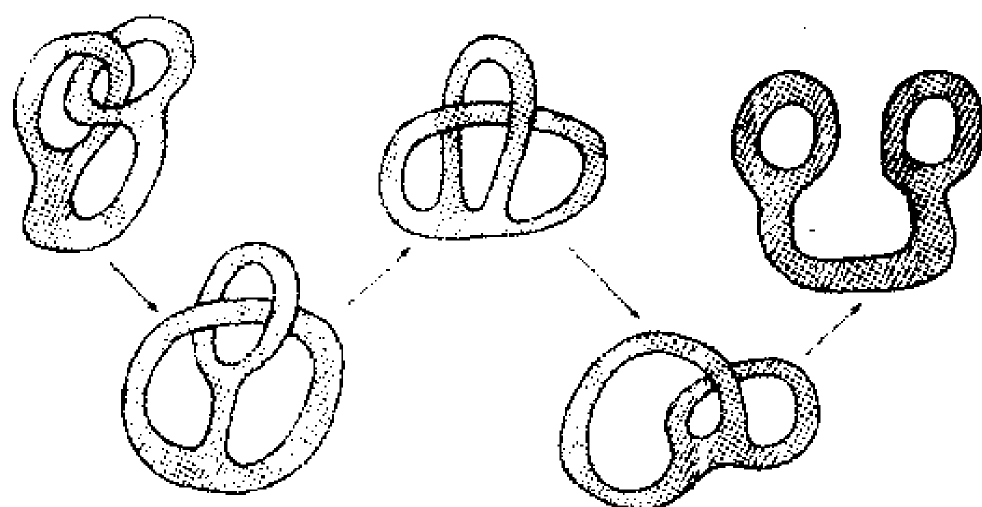


图 1C

在下一步的练习里，曲面上将戳出一个临时性的小洞．我们不仅要把自行车内胎翻出来，而且还想看一看更复杂的曲面，在其内部被翻出来之后将是什么样子．人们马上就会明白，图 2A 那种“中等复杂”的曲面将会变得极其复杂．为了叙述的简化，我们将把内部绘成黑色，外部绘以灰色．我们将在曲面上暂时戳一小孔并标记上洞的“边缘”以作为变形的记录．

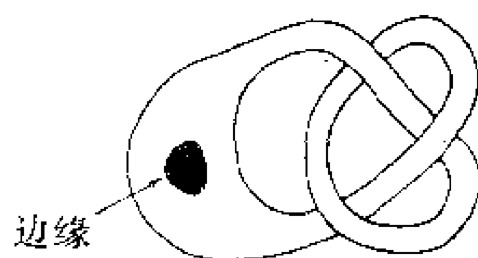


图 2A

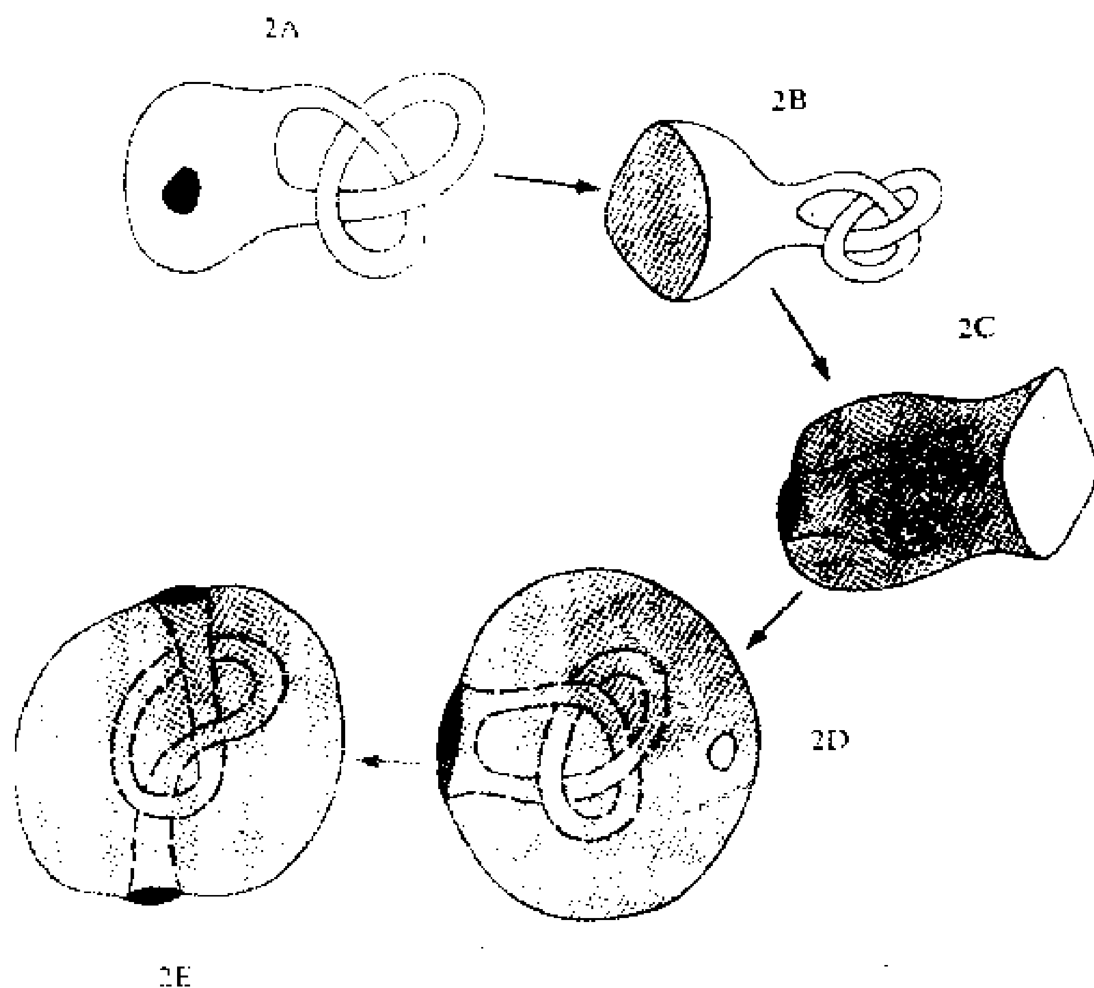


图 2B

图 2B 给出了一系列步骤，藉以表明怎样把图 2A 这个曲面的“里子”翻出来.请注意变形的诀窍：先把复杂的那一部分缩小，再使它穿过此洞.

当洞孔闭合以后，曲面的外侧将全是黑色，而它以前曾是覆盖全部内侧的颜色.用这种画法来说明把曲面内侧翻出来的办法，其优点

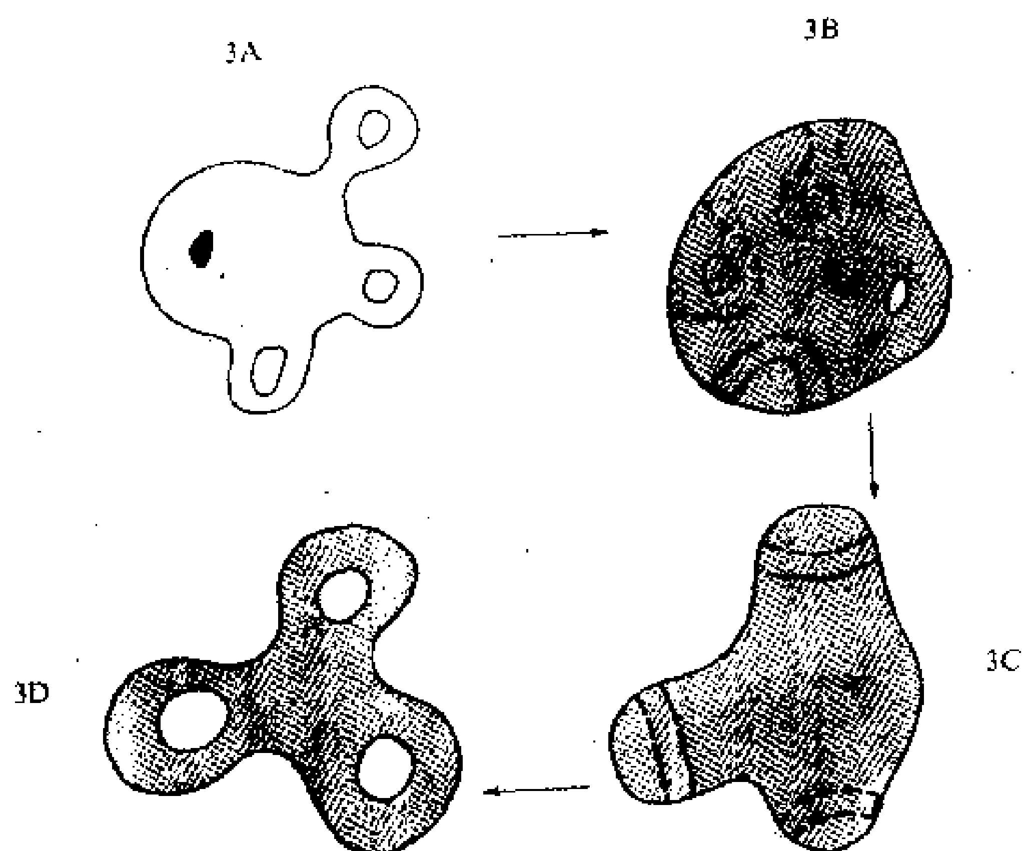


图 3

是很明显的，这在图 3 中可以看清楚. 此法将使一个有好几只“柄”的球在变形后的形状看得清清楚楚，如同人们可以弄清楚自行车内胎的变化一样.

最后的系列变形图将说明图 4 那种套在一起的两个曲面，当其中之一的“里子”被翻出来时，将变成何种模样.

这类图解办法不算新鲜. 对我来说，25 年前已有了这种想法，对某些人也许更要早些. 最近，对纽结理论造诣很深的狄尼斯 · L · 约翰逊(Dennis L. Johnson)提出了一种扭转曲面(见图 5). 现在要请读者做一个轻松的练习，像把图 1A 变换成图 1B 那样，看看你能不能把图 5A 变换成图 5B.

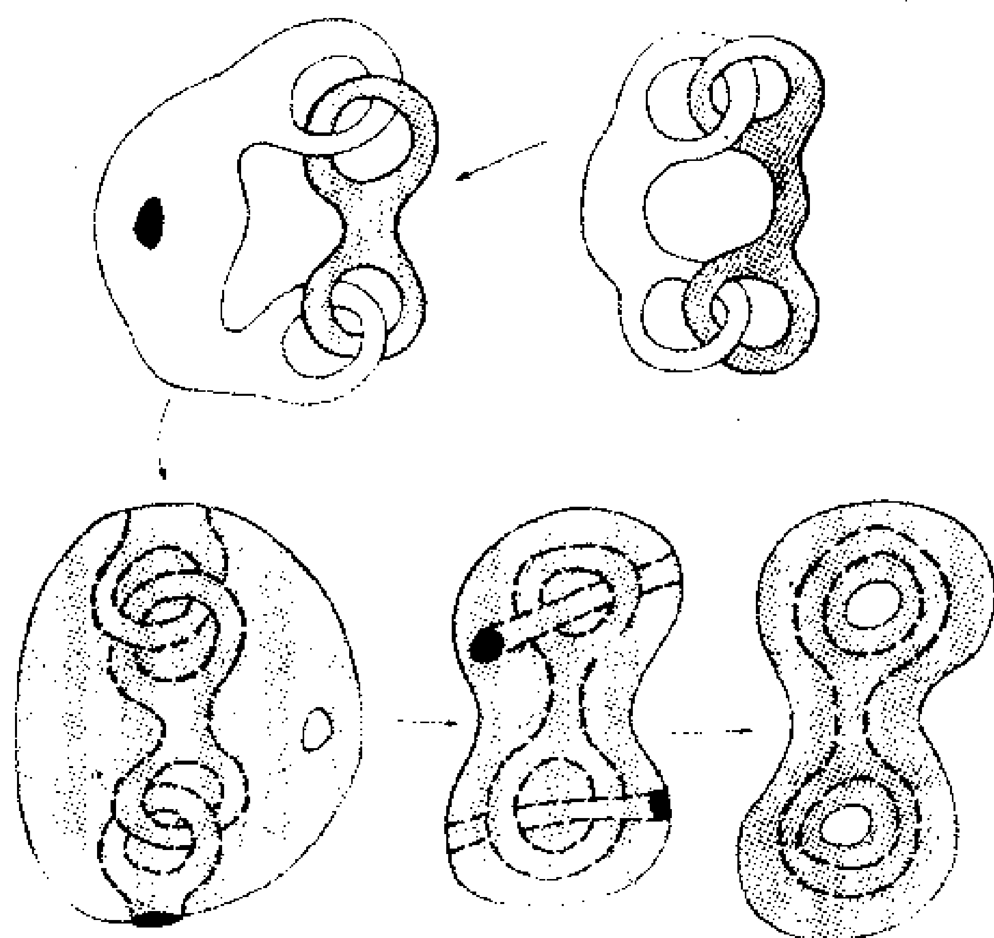


图 4

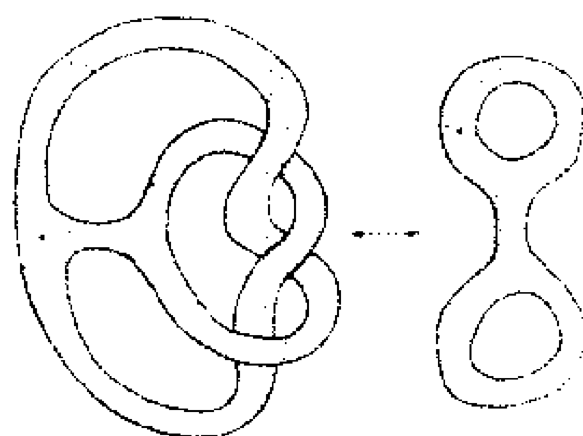
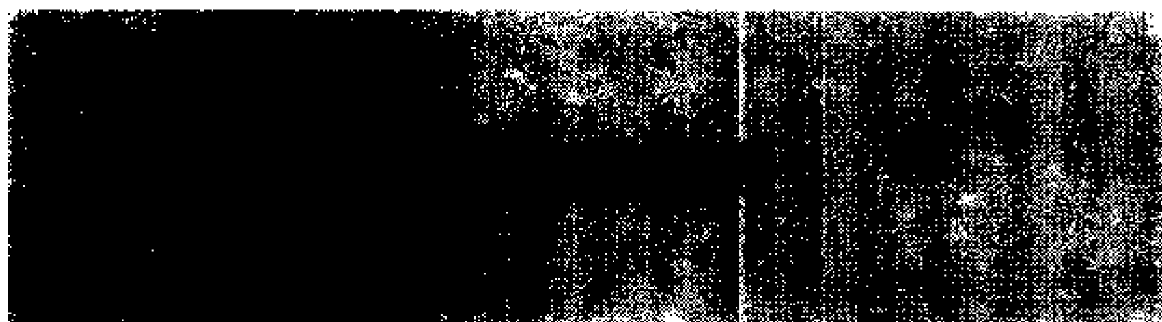


图 5A　　　　图 5B



● 康奈尔大学

□ 罗伯特·康纳利(Robert Connelly)

设有一个封闭多面体，其表面是由几片扁平硬卡纸在边上粘结而成.这类曲面能否弯曲自如呢？也就是说，在不剥掉粘胶纸或扭弯硬卡纸的情况下，曲面能否连续改变它的形状？作为一个实例，让我们考察图1所示的八面体.如果人们用硬卡纸做出这个模型，它将显得非常僵硬，不能弯曲.然而，如果顶部做得比底部略为小一点，它就会砰的一声落下去，如图2，而要做到这一点，人们必须把硬卡纸扭弯.不经畸变，图1是不可能连续地变形为图2的.

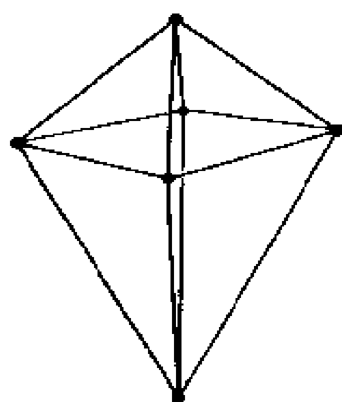


图 1

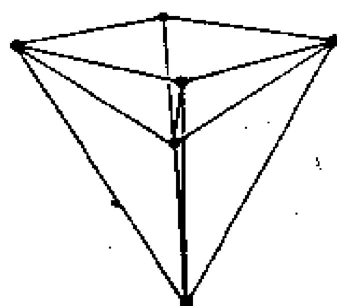


图 2

1813年，著名法国数学家柯西(Cauchy)证明了任意凸多面体的

表面都是刚性的.(不言而喻,所有扁平的自然表面都是紧绷绷的硬卡纸所制成.)于是人们就很自然地猜想,一切曲面,不论其凹凸如何,大概都是刚性的.不幸得很,这一“刚性猜想”是错误的.存在着一个不自身相交的嵌入多面体,其表面可以弯曲自如.下文我将要描述一些由我发现的,并继而由其他学者予以修正的关于否定该猜想的若干实例.

曲面的作法

为了使读者理解我们即将描述的曲面何以能够弯曲自如,我们首先介绍一些可弯曲的八面体,它们是由法国工程师 R·布列卡德(R. Bricard)在 1897 年发现的.这些曲面确实存在着自身相交点,因此我们把它们看作是一些不可压缩,不能伸展的杆棒的组合,在它们的端点用柔韧的橡皮节点予以连接.为了建造这些八面体框架,我们先从图 3 的斜四边形 $aba'b'$ 开始,在此四边形中,对边具有相等的长度.于是可以看出在三维空间中存在着一条直线 L,以使得该四边形关于直线 L 对称.换言之,若四边形绕着 L 转动 180°,它将转变为自身.我们把 $aba'b'$ 看作是八面体框架的赤道,任意选取不在对称直线 L 上的一点 c,用棒杆把 c 与每一个节点 a, b, a', b' 联接起来.不难检定这样的框架确可弯曲自如.我们于是把这种弯曲的构形绕 L 转动 180°,此时 c 点转到了 c' 点,再把该弯曲构形与其合同框架 c'($a'b'ab$)相连.这样,其合体便是布列卡德所发明的八面体之一种,它可说是易于制作的.完成后的全部框架见图 4 所示.请注意,如果把所有的三角形都“装填”进去,则所得之曲面将多次自身相交.我们的目的则是尽可能减少与简化这种自身相交现象.

在图 5 中我们来看另一种形式,开始时,所有的棒杆都落在一个平面上.对称直线 L 垂直于该平面,当骨架扭转运动时,各顶点并不保持在平面之上,然而,在开始时,这样的位置是方便的.

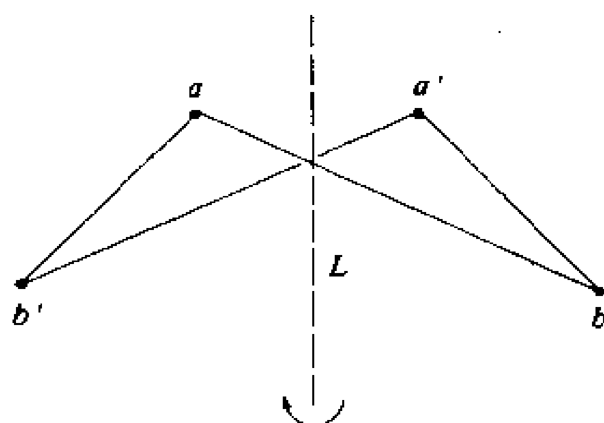


图 3

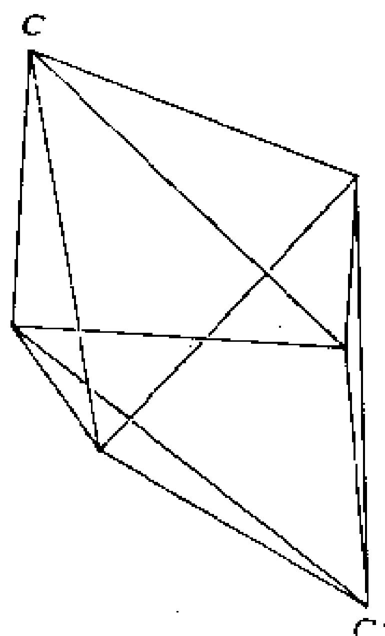


图 4

另一种稍有不同的框架如下，开始时，点 a 与 a' 位于图 5 的水平平面 H 之上. 然后，选取在平面 H 上高度为 $\varepsilon>0$ 的点 b，b' 以及 H 上高度为 $\delta>\varepsilon$ 的另外两点 c，c'，要使得所有这些点在 H 上的正投影正好构成图 5 的形状. 直线 L 仍然垂直于 H，这一八面体框架依然可以弯扭自如，而三角形 $ab'c$ 与 $a'bc'$ 是连在一起的，也就是说，除非强行断裂，它们是拉不开的.

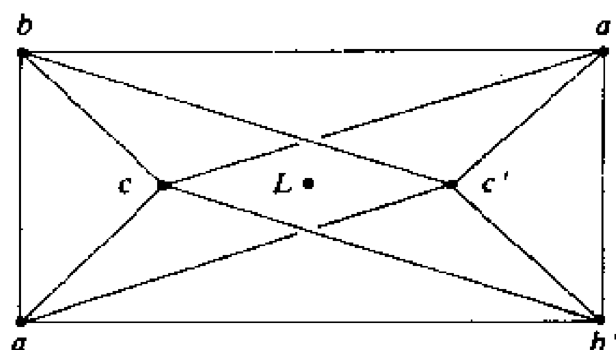


图 5

为了装填柔韧表面，我们可从图 5 开始. 我们并不是要把图中所有的三角形都用扁平的硬卡纸来充填，而是要把曲面稍为改变一下，但是依然要把旧框架中的棒杆继续保持为曲面

的棱边.我们把八面体表面看作是由两片东西做成的——一个底与一个顶.底的形状见图6所示,然后把每个三角形表面往下推,从而得到一个类似于图7的新表面,此时,每个三角形都已被一个顶朝下,又无底的四面体(我们称之为一个凹陷)所取代.



图 6

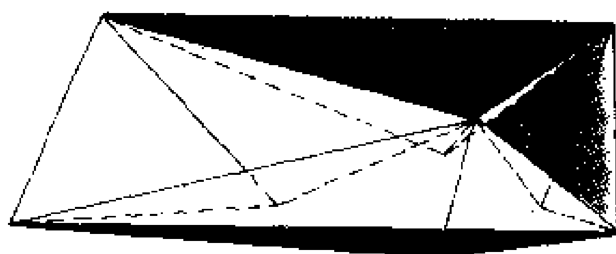


图 7

类似地,我们用图9的那种表面来取代顶面(图8),这时,每一个三角形都已被一个无底棱所取代.正如图6与图8的表面能屈伸自如一样,图7与图9的表面也是如此,自然在后一种情况下,还得加上外加的顶点,也就是说,棱锥的顶点对其底面可作刚性运动.

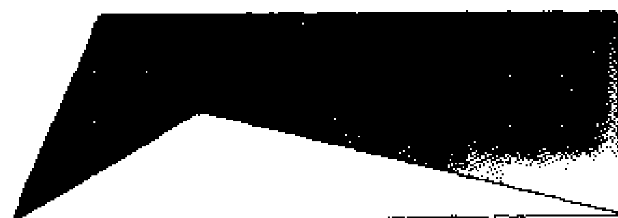


图 8

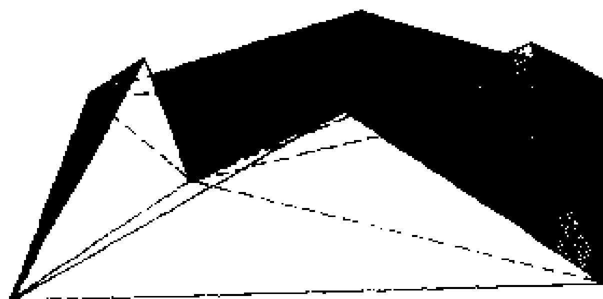
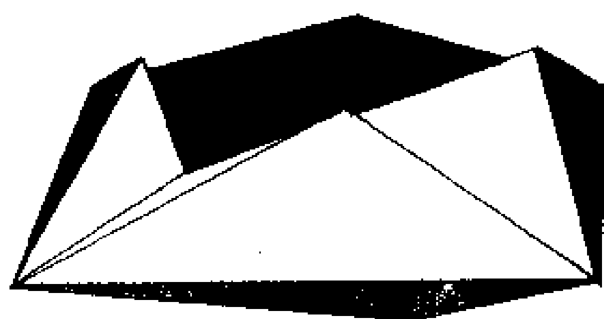


图 9

我们接着把图7与图9沿其共同边界粘贴起来,得出图10那种曲面.它也是扭曲自如,恰如图5的曲面.然而,不幸的是,它有一对自身


图 10

相交点——s与s'.图11给出了图10中相交的这部分表面——点s与s'相当于图5中的交叉点.

为了排除s与s'，我们需要制作一个我称之为褶曲的构形，它同

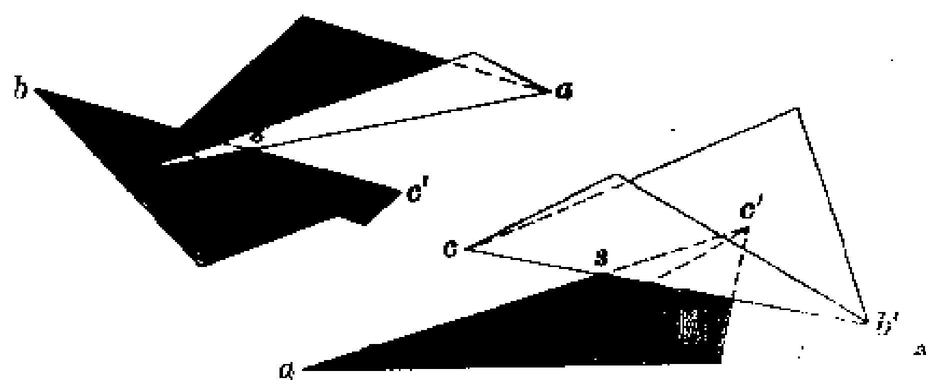


图 11

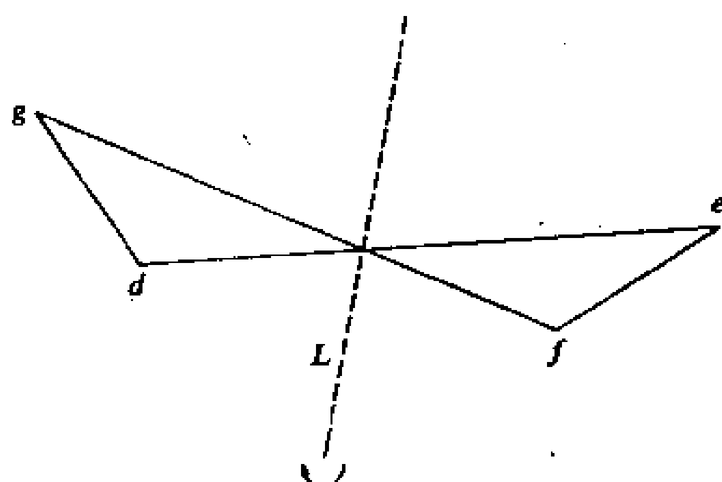


图 12

样由布列卡德(Bricard)的柔韧八面体演变而来. 选取一个对边相等($de=fg, ef=gd$)的平面四边形 $defg$ 如图 12 所示,使线段 de 与 fg 相交. 在通过 $defg$ 圆的圆心之上取一点 h,并以同样的距离在圆心之下取一点 h'. 于是 $hd=he=hf=hg=h'd=h'e=h'f=h'g$. 这样一来,框架 $h(defg)$ 与 $h'(defg)$ 在连接处是牵引自如的.(四边形 $defg$ 实际上一直保持共面.)三角形表面 $hef, hfg, hgd, h'ef, h'fg, h'gd$ 的和集便是褶曲,它是一个八面体除掉了两个三角形表面(见图 13),其边界为 $hdh'e$. 在牵引过程中,从 d 到 e 的距离保持不变. 图 14 告诉我们褶曲的具

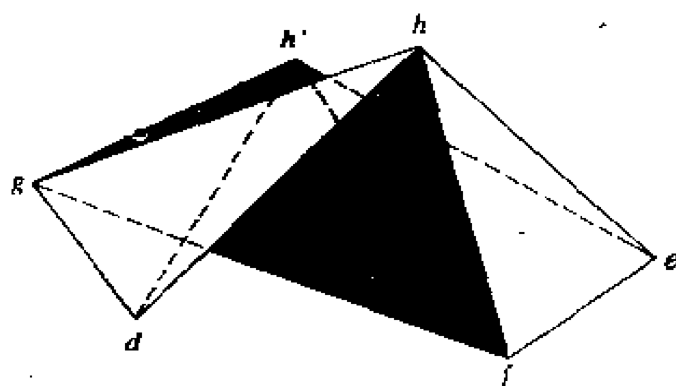


图 13

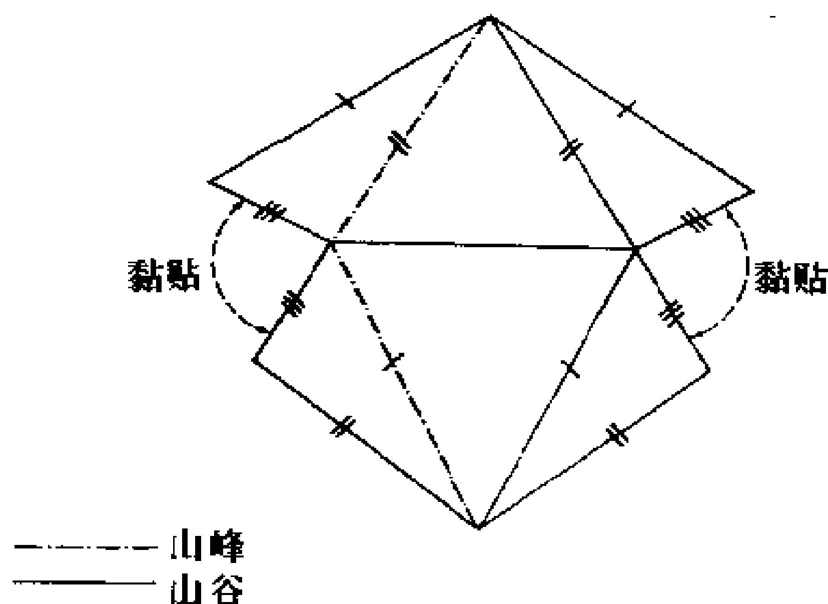


图 14

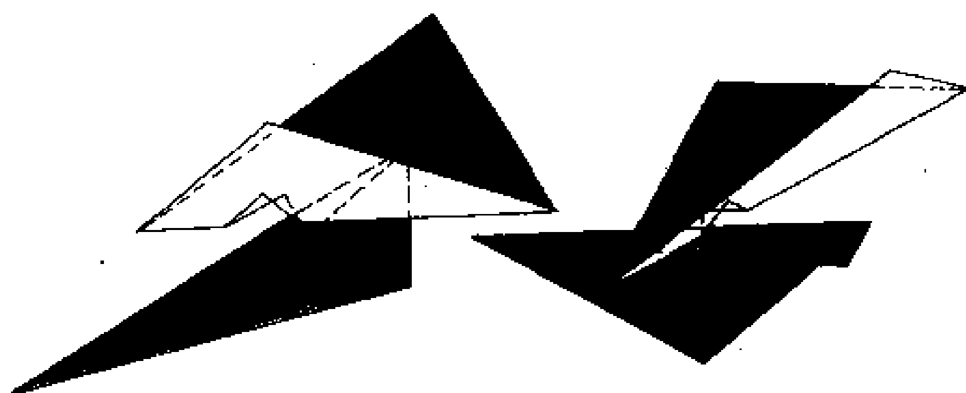


图 15

体制作法.

为了完成柔韧牵引曲面的最后装填工序，可在图 10 的曲面上像图 15 那样，在每个自身相交点的周围切割出一个小洞. 然后在每个小洞中插入一个适当大小的褶曲. 如果褶曲位置安排得恰当的话，则所得之曲面将不会有自身相交点(见图 16). 由于 *de* 在褶曲中保持一固定位置，具有褶曲与两个洞孔的曲面将在接合处屈伸自如，从而整个褶曲面是可以扭拉牵引的，它的外形有点像图 17.

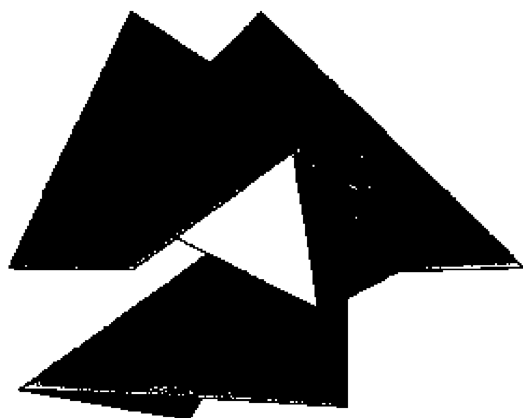


图 16

这个曲面是我所发现的第一个牵引曲面，但我并未注意到其结构能否进一步简化，所需之顶点是否可以减少一些. 后来 N・H・科伊柏(N. H. Kuiper)与皮埃尔・狄里尼(Pierre Deligne)改进了我的设计，得出了一个具有 11 个顶点与 18 个面的曲面. 他们从接在图 5 后面的一段说明中所描述的

框架着手，但却不是在底面上添加四个凹陷，而是只添加了一个(见图 18)，另外三个三角形依然保持扁平状态. 对上部曲面，只添加了两座山峰(见图 19，从两个角



图 17

度给出了它的视图). 当这些新的上、下曲面沿其共同边界粘贴起来后，线段 $c'b$ 将与两山峰的各边在 ca 之上相交. 由于 c, c', b, b' 的略微抬高，这是曲面与其自身

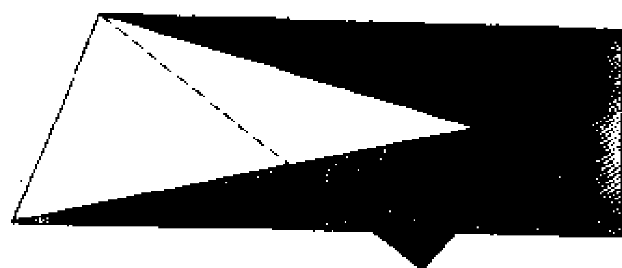
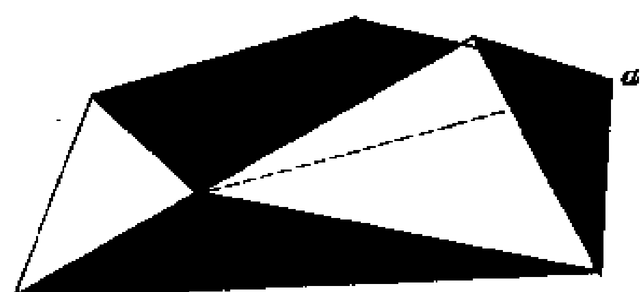


图 18

相交的唯一场所. 他们继而又除去 ca 以及以 ca 为边的两个三角形，并在所产生的洞孔中放入适当大小的褶曲，如图 13 所示. 图 13 中的 d、e 分别与 c、a 贴合，而两个山峰之尖顶为 h 与 h'，图 20 给出了两个视图，表明将上、下部曲面粘贴起来，并从上部曲面除去 ca 后所得之物体形状. 图 21 则从两个角度显示了加入褶曲后最后所得之柔韧曲面形象.

最出色的是，克劳斯·斯蒂芬(Klaus Steffen)发现了一个仅有 9 个顶点的柔韧牵引曲面，他是从类似于图 14 的两个相同的褶曲出发的. 它们与其他两个三角形连在一起，如图 22，这时，图形关于一条铅直线对称，因而各对应长度相等. 其结果是个牵引自如的曲面，有点像图 23.



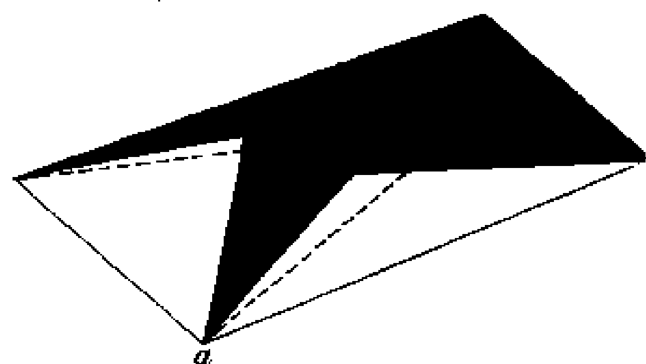
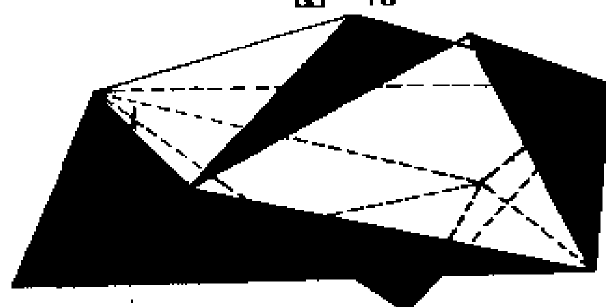


图 19



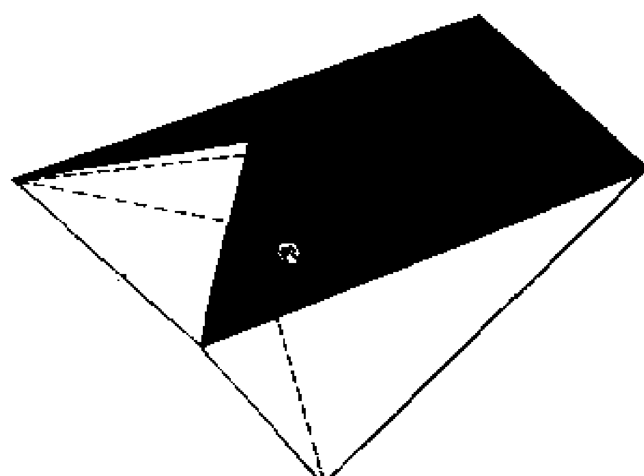


图 20

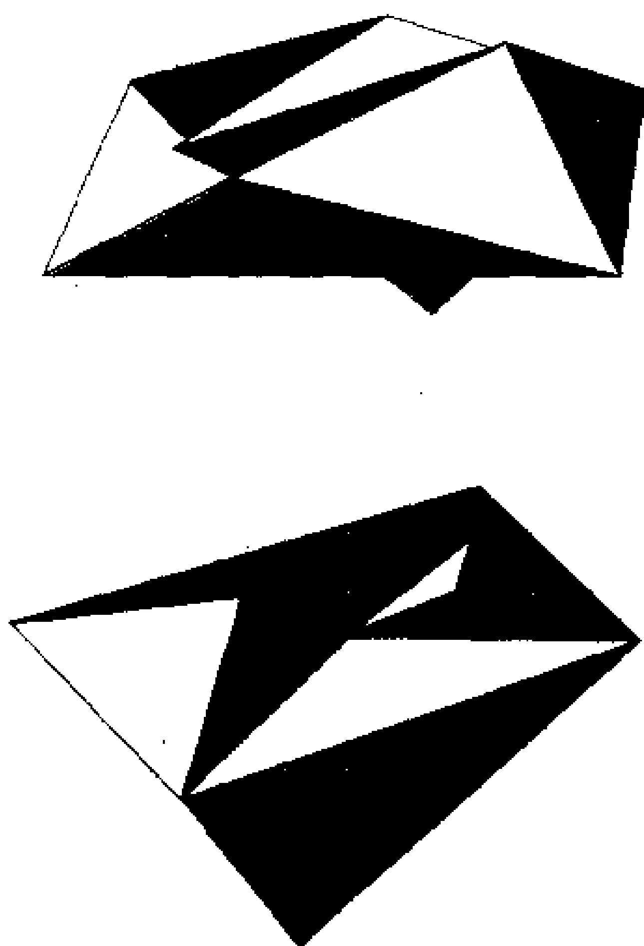


图 21

一 些 猜 想

上述各例子的一个有趣性质是：在把它们牵拉时，被这些曲面包围起来的体积始终保持一定. 虽然如此，我却不知道怎样才能证明对任一可能的牵引曲面，体积都保持一定这件事.

猜想 1 如果一个三角形化的多面体表面是屈伸自如的，则在牵引过程中体积将保持一定.

甚至更为令人惊讶的事情也可能为真. 设 P 与 P' 是三维空间中的两个多面体. 如果我们得以把 P 分割为个数有限的多面体组件 $P_1, P_2, \cdots, P_k$，并通过重新组合而得到 P'（即 $P = P_1 \cup \cdots \cup P_k, P_i \cap P_j \subset$

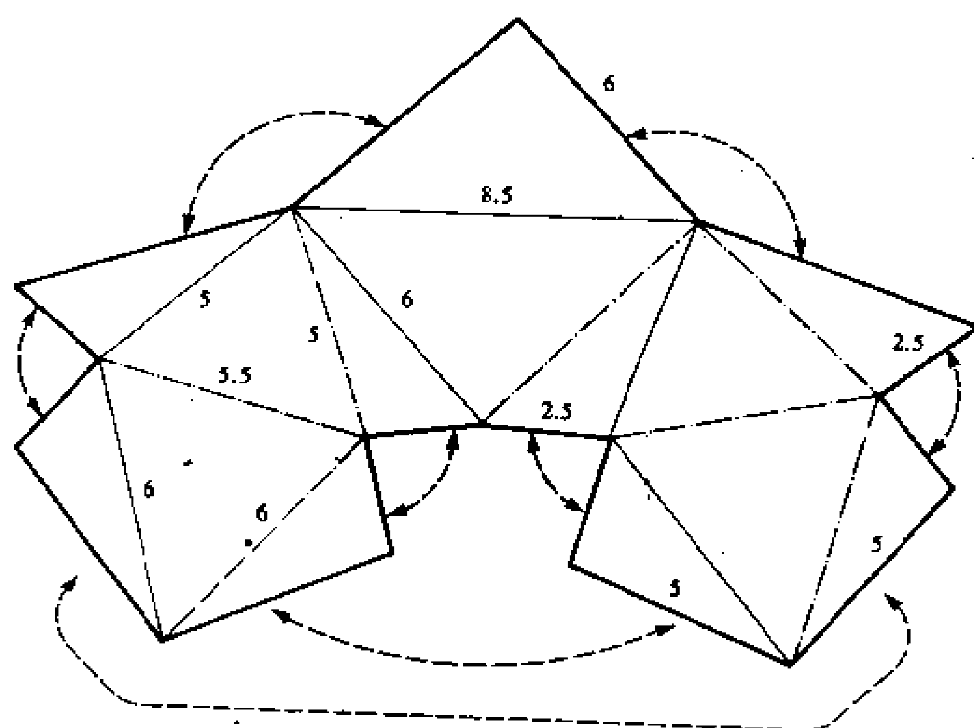


图 22

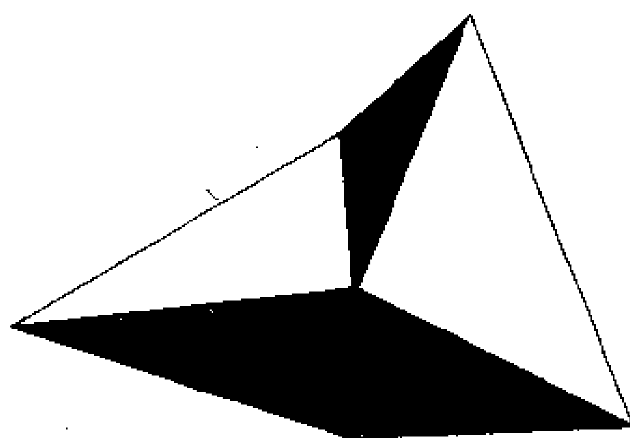


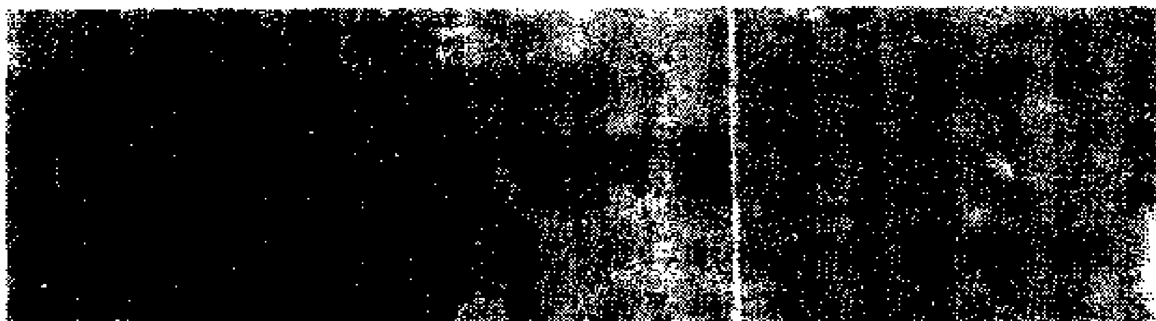
图 23

(境界 P_i)$\cap$(境界 P_j),$i\neq j$,$P'=P'_1\cup\cdots\cup P'_k$,$P'_i\cap P'_j\subset$(境界 P'_i)$\cap$(境界 P'_j),$i\neq j$,对 $i=l,\cdots,k$ 有 P_i 合同于 P'_i),则我们就称 P 与 P' 是分割等价的,写作 $P\sim P'$. 对于这个被希尔伯特(Hilbert)提出的问题 M・迪恩(M. Dehn)的回答是:他发现了具有相同体积的正则立方体与正四面体不是分割等价的.(请参看马丁・加德纳的著作《数学趣题与游戏第二集》第 35 页.)虽然如此,假设 P_t 是由上述柔韧曲面之一在时刻 t 所围成的三维立体,J・P・锡特勒(J. P. Sydler)得到的一个结果表明对牵拉时间区段中的一切时刻 t,将都有 $P_0\sim P_t$. 有关希尔伯特第三问题的一些很有意思的讨论以及锡特勒的这个并非肤浅的结果都可在新近翻译出来的波特扬斯基(Boltianskii)著作中找到(Boltianskii, V. 1978 年出版,《希尔伯特第三问题》. 纽约:John Wiley and Sons 出版公司发行). 当然,一般问题依然悬而未决.

猜想 2 如果 P_t 是被任何柔韧的多面体曲面在时刻 t 包围起来的多面体,则对一切 t 都有 $P_0\sim P_t$.

即使对上文所述的曲面能获悉某种特殊分割法也将是极其有趣的.

原注:本文系根据由美国国家科学基金会部分资助,批件号码为 MCS－7902521 的一项研究工作而写成的.



● 纽约市立大学

□ 斯蒂芬·伯尔(Stephan Burr)

要种九棵树，
横斜得十行，
每行须三棵，
有何千金方？
赐我一妙法，
永志弗敢忘.
更无麻烦事，
把您脑筋伤.

1821 年，约翰·杰克逊(John Jackson)在一本名为《冬天傍晚的推理娱乐》[4]的问题选集中提出了上述数学谜题. 目前，韵文的使用已经不大普遍，一位现代趣题征集者甚至可以干脆不用"种树"这种提法，而把它的说法改变为：在平面上应如何布置九个点，使之成十行，而每行有三点. 当一位数学家遇到此类问题时，他会感到要加以推广的自然动力，因而需要把问题提得更为确切. 这就导致下述提法：给定一个正整数 p，在平面上应如何布置 p 个点($p\geqslant3$)，使得任意四点均不共线，并要使得有三点在一直线上的直线条数为最大？我们将把这一直线的最大条数记为 $l(p)$.

令人生畏的数学家J. J. 西尔维斯脱(J. J. Sylvester)在十九世纪一直在追踪这个难以捉摸的$l(p)$,嗣后它也零星地吸收了人们的注意力,其中既有业余爱好者,也有专业人员.以往,业余爱好者们往往也能对此类问题作出有价值的贡献.可是,目前业余数学家已濒临绝种,这委实是一种羞耻.之所以如此,恐怕一部分原因是一大批数学分支的难以接近性.话虽如此,也还是有许多领域,特别是与组合数学有关的那些分支,业余爱好者依然有用武之地.不幸的是,广大公众对这些易于接近的问题所知不多.通常,他们听到的是诸如费尔马最后定理那样的有魅力的大问题,而对于这些东西,即便是一位专家也难望取得重大进展.因而,组合几何学(这个果园问题是其中一例)的一个引人之点就是:业余爱好者们可对之作出相当贡献.

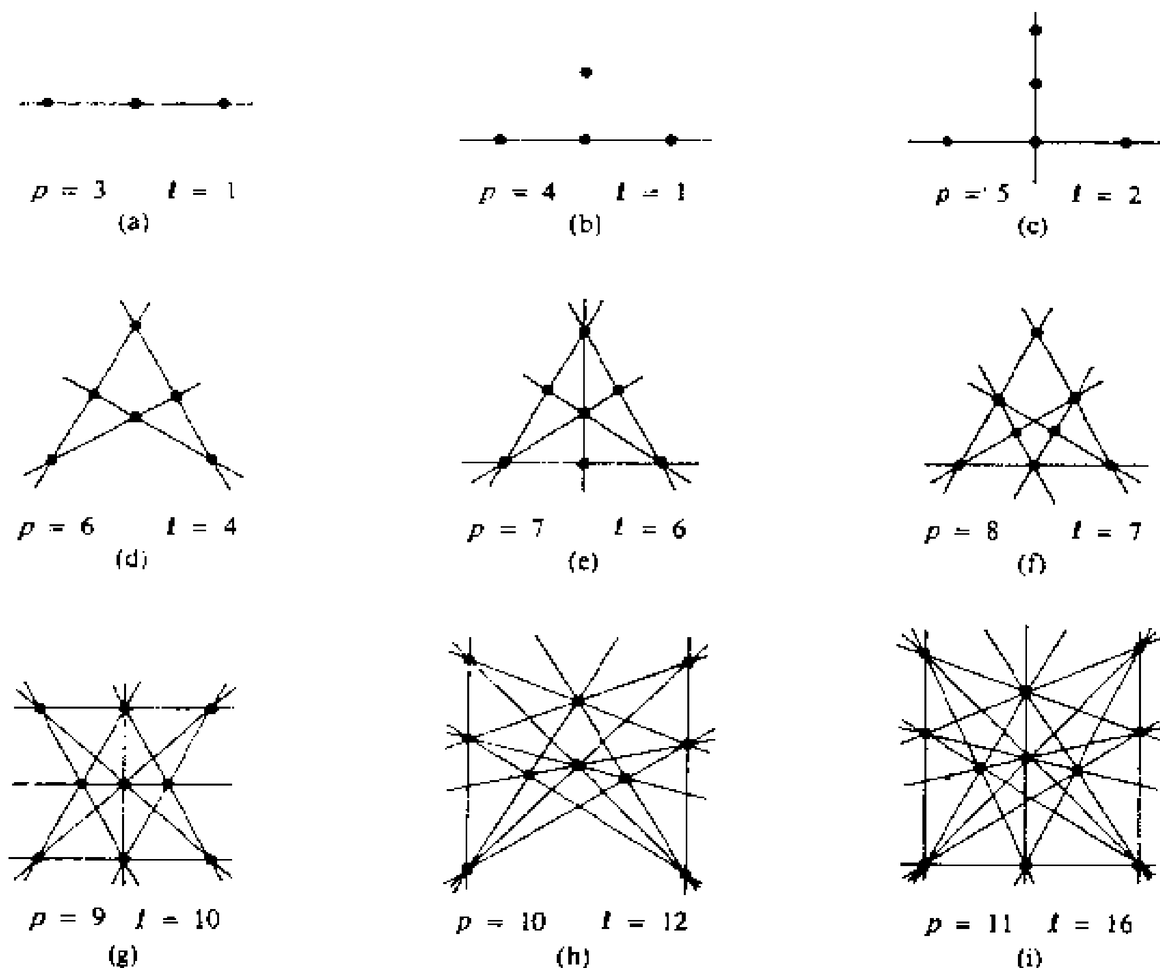


图 1

在图 1 中，给出了相应于 $p=3,4,\cdots,11$ 的一些果园. 所有这些都是最优解，行的个数达到了 $l(p)$. 此外，只有另外两个 $l(p)$ 值是确切知道的，我们马上就要谈到它们. 但是，在目前，请注意，四个点并不比三个点更好些；有时，一个果园图完全被包含在另一个之中，例如 $p=10$ 与 11 的情况.

一个很自然的问题是：上面的安排是否唯一解？回答是否定的. 图 2 给出了 $p=8$ 的另一种不同安排，它是从 $p=7$ 的那个图形导出的，当然那个外加上去的点可位于新直线上的任何地方.

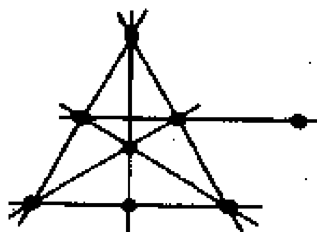


图 2

然而，另有一种方式来表明这些安排不是唯一的：对之可作射影变换. 我们不想完整地解释什么是射影变换（有时也叫射影），但其中的一种类型可照字面加以想象. 把书的一页翻起来，倾斜地看那些图形，由于透视的改变，距离与角度都随之而变，可是直线依旧是直线，因此，变换后的配置仍然具有我们所需的性质.

在作变换时产生了一个问题：如同铁轨那样，平行线将变成不平行，于是使得某一种配置在本质上发生了变化，反之亦然. 然而，数学家堪称是目前一种时髦概念“把问题变为机遇”的发明者，很早以前，他们就对射影变换作了这种处置. 他们在通常的平面上添加了一条包含无穷远点在内的、理想的无穷远直线. 一组平行线被认为相交于无穷远直线上的某点. 如果一个人沿正好相反的方向走，以至于无穷，则认为他所到达的是同一点，就这个意义说，一组平行线恰好有一个交点. 于是现在可以认为任意两条直线必相交于一点的命题正确无误. 于是产生了所谓的欧几里得射影平面与射影几何学，后者已

表明自己是趣味数学的一个重要源泉.

在我们现在的问题中,在简化某些复杂图形并使之更为对称方面,无穷远点将极其有用.它也有助于首先发现那些符合条件的配置,并证明那些我们后面将要提到的结果.作为"安上"无穷远点的实例,让我们在 $p=9$ 的配置中,使顶上面一线的点变为无穷远点.这样就得出图 3.

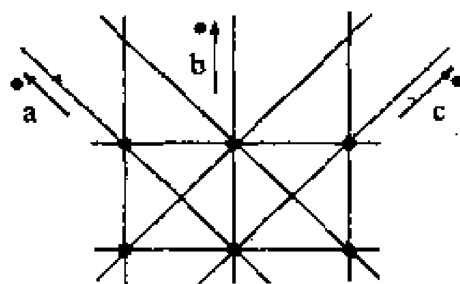


图 3

图上被标记为 a、b、c 的箭头指明了配置中三个无穷远点的动向,任何一个箭头如果变为相反方向也同样可以.当然,无穷远直线也被看作配置中的一行.

图 4 给出了已知为最优配置的、仅有的另外两个果园,即 $p=12$ 与 $p=16$.请注意,$p=16$ 的果园中包含了一个最优的 7 点果园.这两个图中都有三个无穷远点与一条无穷远直线.自然,这两个图都可以进行适当的射影变换,使无穷远点与无穷远直线都变作寻常点与寻常直线,然而,美丽的对称性将会丧失,在一张较小的纸头上也将难于作图.

对 p 的其他值又将如何?表 1 给出了对 $p=3,4,\cdots,25$,有关 $l(p)$ 的一些情况,其中也有最好的已知下界与上界.$l(p)$ 已被明确肯定的十二种情况用星号予以标明并省略了相应的上界.表 1 与本文所谈到的几乎所有成果都引自我同 B. Grünbaum 及 N. J. A. Sloane 合写的、名为"果园问题"[1]的一篇论文.虽然,其中许多结果又是从前人的工作中转引而来.



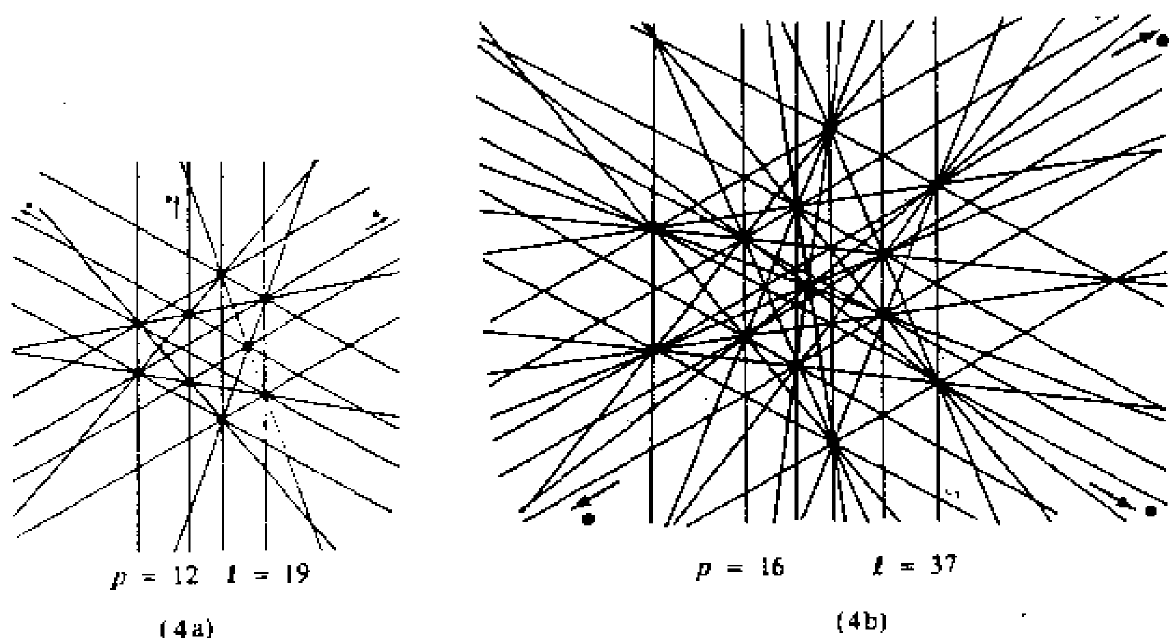


图 4

从 $p=20$ 起，表 1 中所给出的所有上、下界都是下面两个一般定理的结果.

定理 1 $l(p) \geqslant \lfloor p(p-3)/6 \rfloor - 1$，此处，$\lfloor x \rfloor$ 表示不大于 x 的最大整数.

定理 2 如果 $p \geqslant 4$，则 $l(p) \leqslant \lfloor (p(p-1)/2 - \lceil 3p/7 \rceil)/3 \rfloor$，此处 $\lceil x \rceil$ 表示不小于 x 的最小整数.（当然，$l(3)=1$.）

我们不去证明定理 1，但可以给出一些迹象来说明它怎样来自三次曲线（满足三次代数方程的曲线）理论，图 5 画出了由方程

$$(x-1)((x+2)^2-3y^2)=8$$

所定义的对称三次曲线，并同时给出了曲线上的十二个点，其中也包括三个无穷远点. 这十二个点的安排同图 4a 所给出的 12 点果园的布局完全一样.

p	$l(p)$的下界	$l(p)$的上界
3	1*	
4	1*	
5	2*	
6	4*	
7	6*	
8	7*	
9	10*	
10	12*	
11	16*	
12	19*	
13	22	24
14	26	27
15	31	32
16	37*	
17	40	42
18	46	48
19	52	54
20	57	60
21	64	67
22	70	73
23	77	81
24	85	88
25	92	96

表 1

为什么利用三次曲线能产生点的良好配置？其秘密在于根据所谓魏尔斯特拉斯(Weierstrass)椭圆函数，某些三次曲线可以由参数表达式来给出．这种表达法给曲线上的每一点对应一个 0 到 360 之间

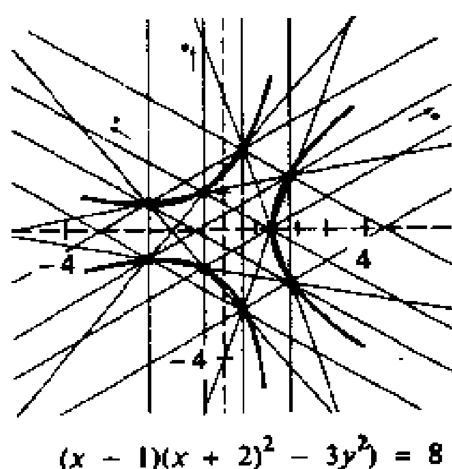


图 5

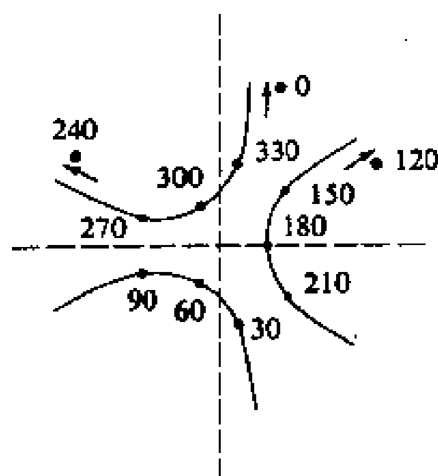


图 6

的实数,当且仅当三个实数之和是 360 的倍数时,相应的三点共线.(只要我们高兴,我们也可以使用其他魔数,但是 360 是便于应用的,让我们不妨把魔数看成是角度之类的东西.)图 6 就是图 5 的变相,仅不过把许多直线抹掉了,而且用参数来表示各点.容易查明图 5 中的每条直线都能满足上述判别准则(包括无穷远直线在内).

对任意的 p,将不难通过选取 p 个在 0 与 360 之间的数,使它们之中的三个数相加之和是 360 的倍数,这样的三数组多多益善.这类选择可以导出一个果园图,从而可以得出 $t(p)$的一个下界.结果便是定理 1 中所说的下界.1868 年,西尔维斯脱[6]证明了一个下界,这个下界除了 p 为 3 的倍数之外,与定理 1 所说的一致.在那种情况下,定理 1 所说的下界要更好一些,但两者的差数也不过是 1.(对声名昭著、犹似西尔维斯脱那类人物的成果,即使只是稍为改进一点点,也将是极为令人鼓舞的.)

除去 $p=7,11,16,19$,表 1 中的任何一个下界都可用定理 1 加以说明.对上述几种情况中的每一种,结果变成可以利用位于一条特殊选定的三次曲线上的 $p-1$ 个点,然后再外加一个不在曲线上的点

所形成的配置，以期得到更优于定理 1 所关联的果园. 作为一个实例，图 7 给出了图 4b 中 16 点果园的作法. 请注意外加的点∅不是 0，因为它不在曲线上，所以它没有相应的参数. 为了使图形看得更清楚一些，只画出了那些通过∅点的直线.

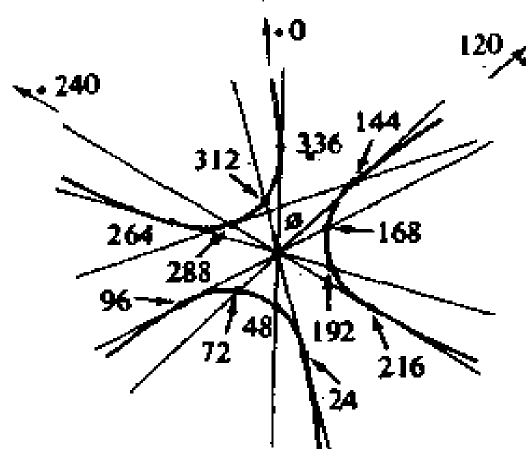


图 7

现在让我们转而讨论 $l(p)$ 的上界，包括定理 2. 所用之手法将大不一样. 一种有用的工具是我们将称之为果园的图的东西. 画出一个果园的各点，然后用一线段来联结两点（此线段称为边），如果它们不与第三点共线的话. 例如图 8 给出了图 1 的 7 点与 8 点果园以及图 2 中另一种 8 点果园的图. 两个 8 点果园的根本不同点在它们的图中看得十分明显，例如，其中之一含有一个三角形，而另一个却没有.

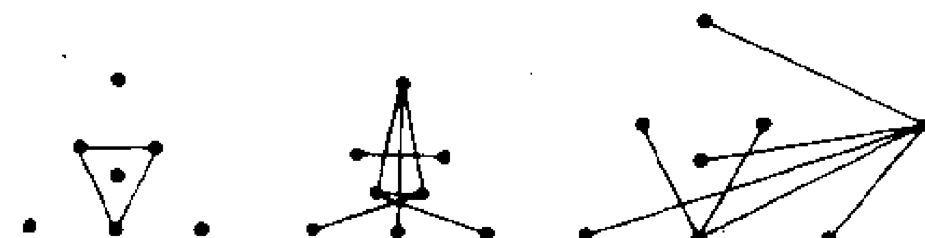


图 8

设果园中有一个点，其中 k 条直线中有三条通过它，则在与之对

应的图中，该点是在 $p-1-2k$ 条边上，这是由于 k 条线将要消去 $2k$ 个点，否则它们将在图中与该点相邻之故.（我们将把这类边的个数称为点的阶.）由于这一事实，故若 p 为偶数，则图中的每个点必为奇数阶，若 p 为奇数，则图中的每一点必为偶数阶（包括零在内），这是由于 p 与 $p-1-2k$ 具有相反的奇偶性之故.

现在进一步考察一个果园的图的边数. 如果一个果园中任何三点都不共线，则每一对点在对应的图上都可给出一条边，于是不难看出这类图（称为完全图）具有 $p(p-1)/2$ 条边. 不过，倘若果园中一旦出现三点共线的情况，图中就将消去三条边.

因此，如果园有 l 条边，则其对应之图将有 $p(p-1)/2-3l$ 条边. 故如 e 为边数，则可得出

$$e=\frac{p(p-1)}{2}-3l,$$

于是
$$l=\left(\frac{p(p-1)}{2}-e\right)\Big/3.$$

由于 $e\geqslant 0$，所以
$$l\leqslant\left\lfloor\frac{p(p-1)}{6}\right\rfloor.$$

但若 p 是偶数，则图上任一点必为奇数阶的，这意味着阶数至少是 1，为此，$e\geqslant p/2$ 必为正确，于是当 p 为偶数时，

$$l\leqslant\left\lfloor\frac{p(p-2)}{6}\right\rfloor.$$

只要稍加变化，上述关于 l 的不等式就可以纳入到下式中去：

$$l(p)\leqslant\left\lfloor\frac{p}{3}\left\lfloor\frac{p-1}{2}\right\rfloor\right\rfloor.$$

这一结果不如定理 2 那样强. 对于那个定理，我们尚需引用一个由 Kelly 与 Moser 发现的定理〔5〕. 该定理断言，在 p 个点的任意配置中，只要并非所有的 p 个点都在同一直线上，那就必然至少存在$\lceil 3p/7\rceil$对点，它们不在一条还有第三点的直线上. 对于一个至少有四点的果园这就意味着在对应的图中，$e\geqslant[3p/7]$，于是根据前面的一段，便有

$$l \leqslant \left\lfloor \left(\frac{p(p-1)}{2} - \left\lceil \frac{3p}{7} \right\rceil \right) \Big/ 3 \right\rfloor,$$

而这实质上就是定理 2.

请注意，在定理 2 的这一证明中并未过分倚重由一个果园推出它的图的机敏作法. 然而，在处理一些特例时，果园的图的概念是非常有用的. 实际上，在上界比定理 2 给出的更好的那些情形中，果园图已表明它们是很有用的. 这些特例是 $p=8,10,12$ 与 14；它们中的每一个情形，由表 1 给出的上界都要比由定理 2 所给出的上界好. 我们将对 $p=8$ 的情形证明一下，至于其他情形，证法也是差不多的，不过要显得冗长一些.

为了证明 $l(8)\leqslant 7$（这同要证明 $l(8)=7$ 实际上是一样的，因为 7 行的果园是确实存在的），我们必须证明不可能存在 8 点、8 行的果园. 我们用反证法，先假定它存在，现在考虑它的图，则其边数 e 将由下式给出，即

$$e = \frac{p(p-1)}{2} - 3l = \frac{8\times 7}{2} - 3\times 8 = 4.$$

此外，由于 8 是偶数，所以图上的每一点都是奇数阶的. 于是，唯一能够满足上述要求的 4 条边的图只能是图 9 那样，具有 4 条孤立边的图.（在图 9 中，所作之图仅是抽象地绘出的，只表明点与点的连接关系而忽略了这些点的真实位置.）

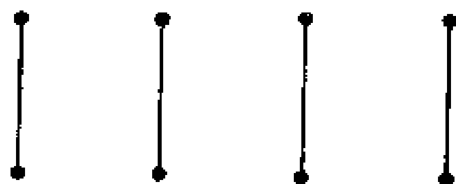


图 9

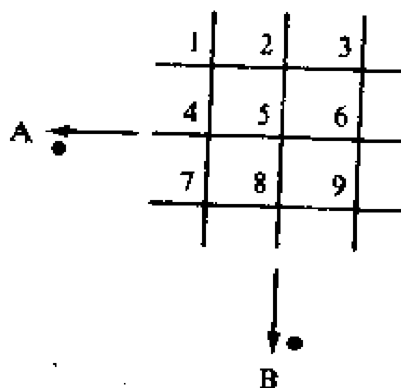


图 10

解答

考察图上的 A 点与 B 点. 转为相应的果园之后，我们可以利用射影变换，把它们放到无穷远直线上去. 对图进行了一番考察之后，我们看到，果园的另外六个点必须落在自 A 发出的三条直线上，也应落在自 B 发出的三条直线上. 这种情况可用图 10 来描述(使图 10 整然有序的唯一办法是等间隔地画出直线).

果园中的六个点(A、B 两点除外)必须落在图 10 所示的九个交点上. 另外，我们只用了(假想中的)果园的六行直线；另外两行直线必须用上这九个交点. 添加新的两行直线的唯一办法是使它们通过点 1,5,9 与 3,5,7—— 但这也不过是用上了七个点，即 A,B,1,3,5,7,9(另外，直线 $A5$ 与 $B5$ 也只有两点落在其上). 于是，待求的果园是不可能存在的，从而 $l(8)$确实等于 7.

关于这个问题，也许可以说上更多的话，对与之有关的问题，甚至可以说得更多. 不过，上面的讨论确已捎来了问题的独特韵味. 我们将给出一些注解，以阐明在本问题方面还有一些什么事情可做，用以结束本章. 显然，在所有事情中最好是能对一切 p，准确地给出$l(p)$的值. 这件事情看起来很困难，但也许可以办得到. 在文献〔1〕中，我们猜测，除了 7,11,16,19 外，对其他一切 p，均有

$$l(p) = \lfloor p(p-3)/6 \rfloor + 1,$$

换言之，定理 1 几乎已经全部讲到了.

人们至少希望，能缩小定理 1 与定理 2 之间的缺口. 如果省略定理中的括号并把两式相减的话，我们可以看到，对每个 p 值来说，缺口大致是 $4/21\ p-1$，可见这个缺口的增长是相当缓慢的，表 1 已证实了这一点.

在任何情况下，决定 $l(p)$的值或者对较小的 p 值设法缩小其缺口都将是极为有趣的. 显然，着手进行探讨的出发点是 $p=13,14$(也有可能是 15). 从上述猜想的观点来看，最好的尝试是设法降低表 1 中的上界. 另外，研究上界问题不一定需要懂得很多三次曲线或诸如此类的知识，而探讨下界问题则有此需要. 那样的探讨将需要不厌其详地进行冗长烦琐的论证. 人们开始时可以用这里已讲过的，用于 p

=8 的那种办法，但是许多不同的分支情况需要分别核查. 有可能一个巧妙的计算机程序在执行研究计划时是有用的.

本文的主旨是想对组合数学的一个有趣角落发射出一些光热. 我也希望鼓舞起有兴趣的爱好者们的信心，在这个问题或类似的问题上一显身手（如想了解本领域中的某些其他问题，请看 Branko Grünbaum 的一本杰作《布置与散开》[3]). 把这类问题编入本书看来是特别合适的，因为在今天，数学科普工作方面干得最出色的人是马丁·加德纳. 事实上，在他负责的专栏[2]中曾经讨论过植树问题的一些方面. 植树问题对一位数学园丁来说，确是一项富有成果的活动项目.

参 考 文 献

1 Burr, S. A.; Grünbaum, B.; and Sloane, N. J. A. 1974. The Orchard Problem. *Geometriae Dedicata* 2: 397-424.

2 Gardner, M. 1976. Mathematical Games. *Scientific American*, 102-109.

3 Grünbaum, B. 1972. *Arrangements and Spreads*. Providence, R. I.: Amer. Math. Soc.

4 Jackson, J. 1821. *Rational Amusement for Winter Evenings*. London: Longman, Hurst, Rees, Orme, and Brown.

5 Kelley, L. M., and Moser, W. O. J. 1958. On the Number of Ordinary Lines Determined by n Points. *Canad. J. Math.* 10: 210-219.

6 Sylvester, J. J. 1886. Problem 2572. *Math Questions from the Educational Times* 45: 127-128.



● 缅因大学

□ 霍华德·伊夫斯(Howard Eves)

六百多年前,西奈(Siena)的圣约翰·哥伦比尼(John Colombini)创建了一个新教派,原来打算是吸收那些善心人入会,他们一心奉献给护理与安葬在腺鼠疫中不幸的罹难者,这场席卷全欧的猖獗瘟疫,几乎吞噬了欧洲人口的三分之一.这个宗教团体名为基督会(它与耶稣会毫无瓜葛,后者在其时尚未建立),在 1367 年取得了教皇乌尔班五世的正式认可.后来,这一教派逐渐衰落,虽然在 1606 年打算复苏该团体的努力曾获得部分成功.一些流言蜚语在背后说坏话,其中显然包括在教堂法所不允许的情况下制造与销售蒸馏酒.这些责难,再加上难于维持一定会员人数,终于导致教皇克利门蒂九世在 1668 年宣布撤销该教派.首尾合计,该教派维持了三百年稍多一些时间.

公元 1613 年,在复活基督会的尝试稍后,一位名叫卡伐利利(Bonaventura Cavalieri)的十五岁意大利男孩被批准为教派的新成员.其后,卡伐利利在该教派的寺院里度完了余生.由于他的这一献身精神,也由于随之而来的教派的解体以及基督会与更有名气的耶稣会在名称方面的类似[❶],因而今天许多主要的百科全书,人名辞典,历史与掌故书等都把卡伐利利错误地当做一名耶稣会会员了.这

❶ 译者注:在原文中,两会的名字分别为 Fesuat 与 Fesuit,仅差一个字母.

件事，为我们提供了一个很好的实例——在我们的历史书中确实隐藏着某些永久性的错误.这类性质的错误并不少见，有的已经持续了很长时间.

卡伐利利在1598年生于意大利的米兰.年轻时他曾就学于伽里略.后来，在1619年他就任波罗那大学数学教授，在那里一直工作到他1647年逝世为止，死时还很年轻，仅49岁.他是当时最有影响的数学家之一，在几何、三角、天文学、占星术与光学方面写了不少著作.纳皮尔(Napier)发明对数，在首批认识其伟大价值的人们中间，他起了突出作用，把它引进意大利.然而，他对数学的最重要贡献是1635年初版刊行的论文《极微几何》.这一著作专门讨论微积分之前的所谓“微分法”，正如许多现代数学成就那样，此种方法可追溯到早期希腊学者，例如德谟克利特(Democritus)(公元前410年在世)、阿基米德(Archimedes)(约公元前287—公元前212)，虽然其直接动力也许来自约翰·开普勒(Johann Kepler)(1571—1630)所作出的积分尝试.无论如何，卡伐利利的著作《极微几何》在1635年的出版是数学史上的一个重要里程碑，特别对于微积分史更是如此.

卡伐利利的伟大论文行文流畅但不明确，从中难以搞清楚卡伐利利所说的“微分”究竟意味着什么.看来似乎给定的一片平面的“微分”是指那片平面的弦，而那片平面可认为由无穷多个相互平行的微分(基本部件)叠加而成.与此类似，一个给定立体的微分是那个立体的截面，该立体可视为由无穷多个相互平行的这类截面(基本部件)叠加而成.卡伐利利又论证如果把某一给定平面片的平行微分沿着它自身的轴逐个都作滑动，使这些微分的端点仍然描绘出一条连续境界曲线，则新的平面切片仍将与原有的平面切片等积，这是因为它们都是由同样的一些微分叠加而成的.类似地，如把某一给定立体的一些平行微分沿其自身平面滑动，假若这些微分的边界依然形成一个连续曲面，则新、旧两立体等积，因为它们也是由同样的微分叠加而成.最后这件事实可用一叠垂直卡片进行演示，把这叠卡片倾斜推出，使其边缘形成曲面，显然斜摆与垂直放置的两叠卡片，其体积完

全相等.

把以上说明略加推广,就得出所谓卡伐利利原理.

1. 一对平行直线之间夹有两个平面切片,如果任意一条平行于包线的直线被两个平面切片截得的两条线段的长度总是成一定比,则这两个平面切片的面积之比也等于该定比.

2. 一对平行平面之间夹着两个立体,如果任意一个平行于包面的平面截两个立体所得到的两个截面面积总是成一定比,则这两个立体的体积之比也等于该定比.

卡伐利利把"微分"视为基本部件的模糊概念常煽起对一个图形的各种讨论,从而研究这一课题的学生纷纷对之提出尖锐的批评,特别是瑞士金饰工匠兼数学家保罗·古尔亭(Paul Guldin)(1577—1642).卡伐利利于是作了重新处理,企图弥合分歧,但是他的新尝试并不比他原来的更为成功.法国几何学家兼物理学家吉尔斯·罗伯瓦尔(Gilles Persone de Roberval)(1602—1675)宣称他在卡伐利利之先即已发明了此种办法,然而优先权之争很难解决,因为罗伯瓦尔对披露他的发明一事素来磨磨蹭蹭,已经卷入了好几起发明优先权之争.

罗伯瓦尔的拖拉作风也有其原因,从1634年开始,他在皇家学院一直当了40年之久的教授.每隔三年,这一职位就自动变为空缺,由退职的前任教授出题,然后由公开数学竞赛中的优胜者继任教授.自然,为了永保职位,他要留一手,有意把他的发现秘而不宣以用作竞赛题目.不过,总的说来,罗伯瓦尔还是很好地利用了微分法,解决了一批面积、体积与重心问题.此法及其相仿的办法也被下列学者卓有成效地运用了,他们是托里拆利(Evangelista Torricelli)(1608—1647),帕斯卡(Blaise Pascal)(1623—1662),费尔马(Pierre de Fermat)(1601?—1665),文森(Grégoire Saint-Vincent)(1584—1667),巴罗(Isaac Barrow)(1630—1677)以及其他学者.

在面积与体积计算中,卡伐利利的两条原理是极有价值的工具,应用现代积分学知识,它们易于做得更严密些.由于应用这两条原理

在直观看来极为自然，因而一些需要用更先进的微积分技巧来处理的许多求积问题都可用它们加以解决. 实际上，许多立体几何初等教材的作者，也以教学为理由拥护这种设想，成功地应用了卡伐利利第二条原理，因为这样做大大地简化了学生的教材. 例如，在推导熟知的四面体体积公式 $V=\frac{Bh}{3}$ 时，困难在于先得证明底、高相等的任意两个四面体必然等积. 从欧几里得的《几何原本》起，一切立体几何课本中的处理方法都碰到这个潜在难点. 但是，只要一旦应用卡伐利利的第二条原理，困难马上就烟消云散了.

现在我们要列举一些卡伐利利原理的应用实例. 对那些喜欢做数学难题的人来说，有的题目着实可以绞尽脑汁. 为了方便起见，我们先来说一个定义. 如能适当放置两个平面切片，使一族平行线中的任一直线在其上都截出等长线段，或者，如能适当放置两个立体，使一族平行平面中的任一平面在其上都截出相等面积，则所述之平面切片与立体称为是卡伐利利合同的. 由卡伐利利原理，两个卡伐利利合同的图形必然具有相等的面积（第一种情形）或相等的体积（第二种情形）.

例　　题

1. 我们来求半轴长分别为 a,b 的椭圆面积. 设椭圆方程为

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\ a>b,$$

圆的方程为

$$x^2+y^2=a^2,$$

在同一直角坐标参照系上作它们的图形，如图 1.

在上述两方程中分别解出 y，得

$$y=\left(\frac{b}{a}\right)(a^2-x^2)^{\frac{1}{2}},y=(a^2-x^2)^{\frac{1}{2}}.$$

可以看出，椭圆及圆上对应的纵坐标之比是 $\frac{b}{a}$. 相应的椭圆与圆的垂直弦弦长之比也等于这一比值. 于是，根据卡伐利利的第一条原

理，椭圆面积与圆面积的比值也应如此，于是我们得出

$$\text{椭圆面积} = \left(\frac{b}{a}\right)(\text{圆面积})$$
$$= \left(\frac{b}{a}\right)(\pi a^2) = \pi ab.$$

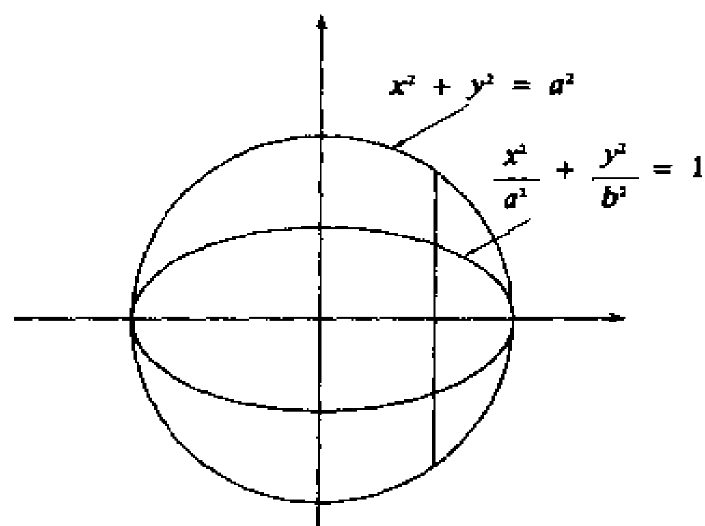


图 1

这与开普勒在求半轴长为 a，b 的椭圆面积时所用的方法本质上是一样的.

2. 现在我们来求熟知的半径为 r 的球体积公式，在图 2 中，左

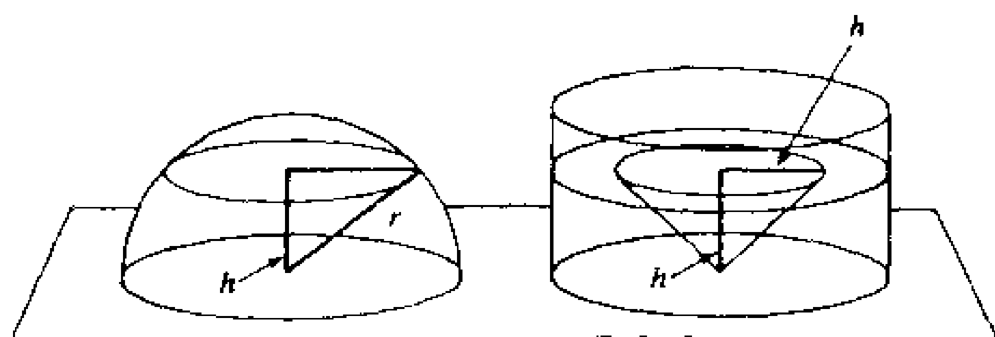


图 2

边是一个半径为 r 的半球，右边是一个半径与高均为 r 的圆柱再除掉一个圆锥后所余的立体图形，该圆锥的底是圆柱的上底，而其顶点则是圆柱下底的中心. 半球与凿空的圆柱都摆放在公共水平平面上. 现在我们用一个位于底面之上、高为 h 的平行平面去截这两个立体，它将分别割出一个圆形与圆环形截面(如图 2)，利用初等几何，立即可以看出这两个截面积都等于 $\pi(r^2-h^2)$. 于是，根据卡伐利利的第二条原理，该两立体必具有相等体积. 于是，球体积可用下法算出

$$V=2(\text{圆柱体积}-\text{圆锥体积})$$
$$=2\left(\pi r^3-\frac{1}{3}\pi r^3\right)=\frac{4}{3}\pi r^3.$$

当然，解题的诀窍在于找出一个"比较的立体"(在此情形，它是一个凿空圆柱)，它是半球的卡伐利利合同体.

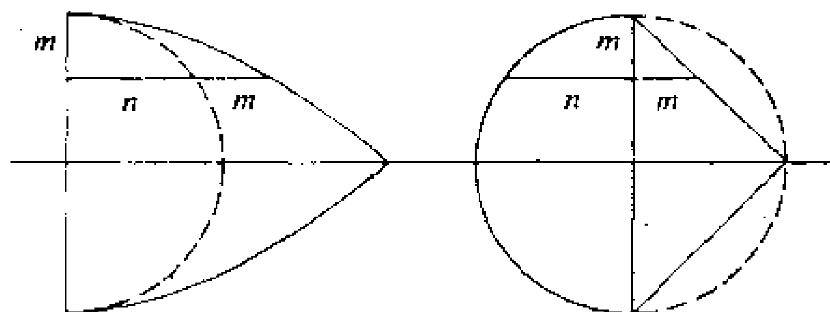


图 3

3. 作为卡伐利利原理平面情形的第二个例子，考察由一条直线与两段弧所围成的平面切片，见图 3 的左面部分，图中注有 m 的两个线段恒为等长. 作为比较用的面积，人们容易得出一个由半圆与等腰直角三角形所围成的平面切片，如图 3 的右半部分. 因而所要求的面积是

$$A=\frac{1}{2}\pi r^2+r^2=\left(\frac{\pi}{2}+1\right)r^2.$$

4. 作为卡伐利利原理立体情形的另一例题，让我们来求球环的

体积，它是从一个半径为 r 的实心球体中凿出一个与球的南北极同轴，且其半径为 a 的圆柱形孔（见图 4 的左面部分）．设有一球，其直

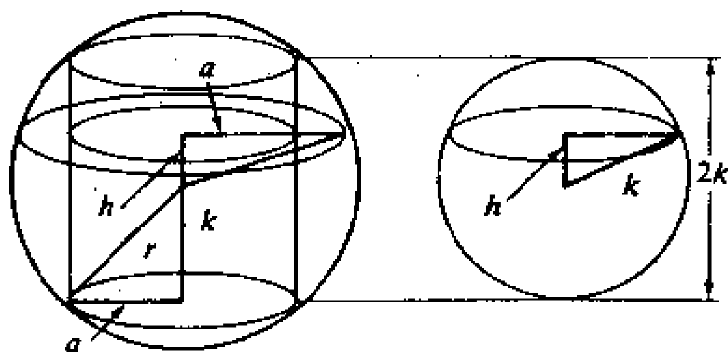


图 4

径等于球环的高，且球心与球环的中心位于同一水平面上（见图 4 的右面部分）．现从距两立体中心为 h 处作一水平面来截两立体．此时，球环的截面为一环状区域，其面积等于

$$\pi(r^2-h^2)-\pi a^2=\pi(r^2-a^2-h^2).$$

而球的截面是一个圆，其面积等于

$$\pi(k^2-h^2)=\pi(r^2-a^2-h^2).$$

于是，由卡伐利利的第二条原理，可知球环体积 V 与半径为 k 的球体积相等，亦即

$$V=\frac{4}{3}\pi k^3.$$

有意思的是凡是有相同高度的所有球环都具有相同的体积而与该环的半径无关．

5．《美国数学月刊》1941 年 3 月号问题 E465 是试求曲线

$$b^2y^2=(b+x^2)(a^2-x^2)$$

所围的面积，这里 $b\geqslant a>0$．该曲线的图形（见图 5）含有一个通过点 $(0,\pm a)$，$(\pm a,0)$ 的非圆形的环，外加一个位于 $(-b,0)$ 的孤立点．此曲线关于 x 轴对称．现考察圆 $x^2+y^2=a^2$，我们将证明所求之面积

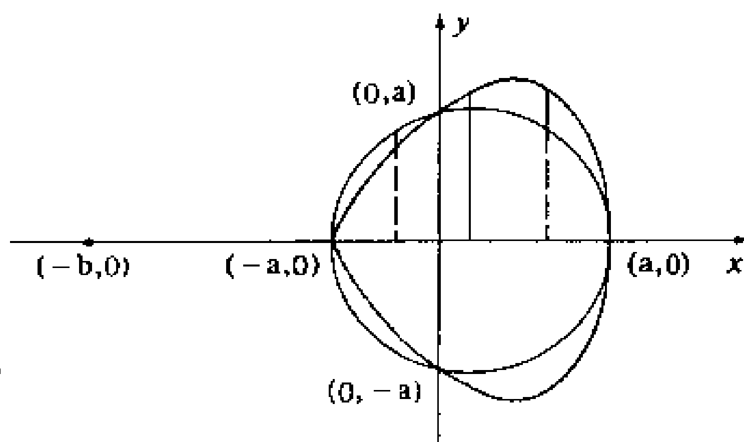


图 5

即等于该圆面积. 由于图形是关于 x 轴对称的，我们只须证明该曲线与圆在第二象限所围成的新月形面积与第一象限的月牙形面积相等就行了. 对这一点而言，我们只须证明两个月牙形中左右横坐标等距离处的相应纵截线段相等. 令 y_1 与 y_2 分别表示同一横坐标 x 处所对应的给定曲线上与圆上的纵坐标，于是我们有

$$y_1 - y_2 = \frac{1}{b}(b+x)\sqrt{a^2-x^2} - \sqrt{a^2-x^2} = \frac{x}{b}\sqrt{a^2-x^2}.$$

可以看出，除符号外，对 $+x$ 与 $-x$，y_1-y_2 的值相同，这就表明两个月牙形区域是卡伐利利合同的，因而被给定曲线所围成的面积等于 πa^2.

问　题

设计一个合适的比较立体，以便利用卡伐利利的第二条原理来计算所需体积，这有时可以构成非常有趣的问题. 也许有读者对这类题目跃跃欲试. 下面举出一些，解法提示可参看本章末节.

6. 利用卡伐利利的第二条原理，试求圆环(锚链)的体积，该物体是把一个半径为 r 的圆绕着一条直线(位于圆所在的平面上，且距圆心的距离为 $c \geqslant r$)回转而成.

7. 不难证明没有一种多边形可以同圆卡伐利利合同(因为在多边形两条边之间,等间隔的弦将均匀地改变其长度,但是在圆上的等间隔弦却并非如此). 人们也会同样认为没有任何一种多面体可同给定的球卡伐利利合同. 请把四面体作为球的比较立体来证明这种想法是错误的.

8 通过正圆柱底面中心的一个倾斜平面把圆柱截出一个楔形,也叫马蹄形(见图 6). 请利用卡伐利利的第二条原理求马蹄形的体积,用有关圆柱的半径 r 与马蹄形的高 h 加以表示.

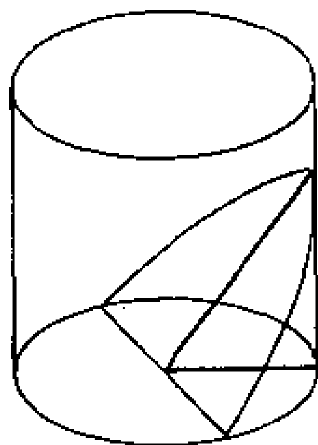


图 6

9. 拟柱是这样一种立体,它有两个平行的底面,且任一平行于底面的平面截此立体所得之截面积是该平面与拟柱一个底的距离的二次函数.

a. 证明棱柱、楔形(转动一个直三角棱柱,将其侧面作为底)及棱锥的体积可由下列拟柱体公式

$$V = \frac{h(U + 4M + L)}{6}$$

给出,此处 h 为高,U,L,M 分别是上底、下底与中截面的面积.

b. 由卡伐利利的第二条原理,证明拟柱的体积可从拟

柱体公式导出.

c. 试证明(1)球(2)椭球(3)马蹄形(见上面的问题 8)(4)斯坦因梅茨(Steinmetz)立体(具有相同半径且其轴互相正交的两个正圆柱的公共部分)都是拟柱体的特例,并由此求出它们体积的表达式.

* * *

下面要用一个告诫与卡伐利利合同的一个惊人性质来结束本章. 先说告诫.

在初等教材里卡伐利利原理常以直觉显然成立而被接受. 现在我们来考察夹在一对平行平面之间的两个立体,而且与包面平行的任一平面所得之两个截面的周长相等. 人们往往认为,根据直觉理由,犹如卡伐利利原理的情形一样,这两个立体的侧面积也是相等的. 之所以得出这一结论的原因是,人们把两个立体的侧面看作是以一系列绝薄细线回路为基本部件(形成立体横截面的周长)累积而成. 由于对应的回路是等长的,而且一个立体的回路与另一立体的回路之间存在着一一对应关系,于是看来两个立体的侧面非相等不可.

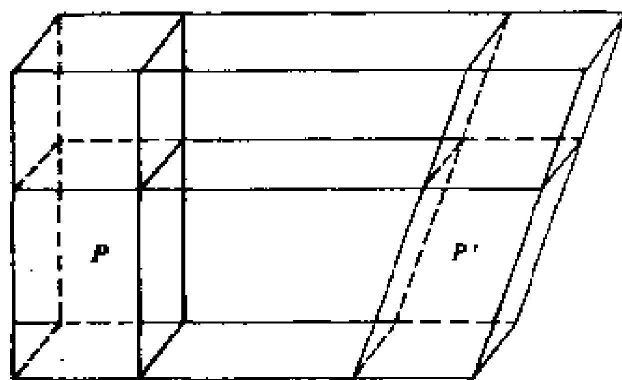


图 7

为了证明此点并非必然如此,可以考察两个正方棱柱 P 与 P',它们的底与高都相等,此处 P 是一个正棱柱,而 P' 是一个斜棱柱,其一对对立面也垂直于底(见图 7). 斜棱柱的侧面积大于正棱柱的侧面积,

这就指出了数学上单纯依赖人的直觉的危险性.

现在让我们考虑卡伐利利合同的惊人性质,让我们先提出一条定义.

同一平面上的两个三角形 ABC 与 $A'B'C'$ 被称为互为仿射映像,如果 AA', BB', CC' 都互相平行,且其中点位于同一直线 m 之上(见图 8),人们得以建立以下两个有关定理.

定理 任意两个共面且面积相等的三角形,不超过三次仿射映像,人们总是可以把其中一个三角形转变为另一个三角形.

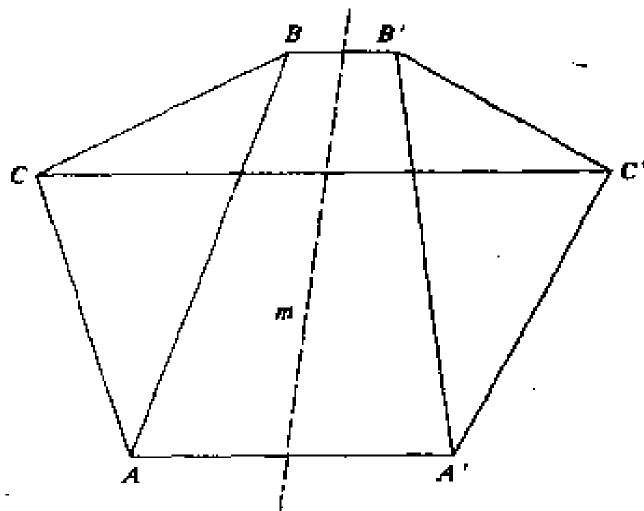


图 8

定理 任意两个等积的三角形必为卡伐利利合同图形.

以上两个惊人定理仅仅是在最近被发现的.第二个定理在三维空间中的类比命题不成立.这就是说,两个等积的四面体不一定是卡伐利利合同的.我们不拟在这里证明这些结果.

问题的解法提示

6. 把圆环置于一个与其轴垂直的平面 P 上,取一个半径为 r,

高为 $2\pi c$ 的正圆柱作为比较立体，并将其横向地置放于平面 P 上，用一个平行于平面 P 的平面来截此圆环与圆柱. 则圆环的截面 A 是一个外径与内径分别为 a,b 的环状区域，而圆柱的截面 A' 则是一个长 $2\pi c$、宽 w 的矩形 A'.

显然

$$A = \pi a^2 - \pi b^2 = \pi(a^2 - b^2) = \pi(a + b)(a - b) = 2\pi c(a - b).$$

而
$$A' = 2\pi cw = 2\pi c(a - b).$$

由于 $A = A'$，由此可知圆环面（锚环）的体积等于圆柱体积，亦即

$$V = \pi r^2(2\pi c) = 2\pi^2 r^2 c.$$

7. 设 AB 及 CD 是满足下列条件的空间两线段：(1) $AB = CD = 2r\sqrt{\pi}$，(2) AB、CD 都垂直于联结其中点的直线，且相距 $2r$，(3) AB 垂直于 CD. 则四面体 $ABCD$ 可视为比较用的多面体.

8. 用一个通过圆柱轴线的平面 P 把马蹄形分为两相等部分. 令 A 为马蹄形的截面的面积，构造一个正棱柱，它的底是一个正方形，面积等于 A（底在平面 P 上），高等于圆柱半径 r. 从此棱柱中截出一个棱锥，其底是棱柱中不在 P 上的那个底，而其顶点则是在棱柱另一个底上的一点. 这一个凿空的棱柱可用作比较立体，以同马蹄形的一半进行对照.

也可以另选一个比较立体，同马蹄形的一半进行对照，其底面是一直角三角形（直角边为 r 及 h），高为 r，而其上底则是平行于下底斜边的一个线段，且与之相等.

马蹄形的体积是 $\frac{2}{3}hr^2$.

9b. 任一截面积，如果是自此截面到一个底面距离的二次函数，则必然是以下几部分的代数和：一个棱柱的固定截面积，一楔形的截面积（与它到底面距离成正比）以及一棱锥的截面积（与它到底面距离的平方成正比），因而拟柱体积等于棱柱、楔形与棱锥体积的代数和，然后再应用(a).

9c. 椭球 $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$

被一平面(与 xy 平面的距离为 z)所截,其截面是椭圆

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1-\frac{z^2}{c^2},$$

其半轴长为

$$\frac{a}{c}\sqrt{c^2-z^2} \text{ 与 } \frac{b}{c}\sqrt{c^2-z^2}.$$

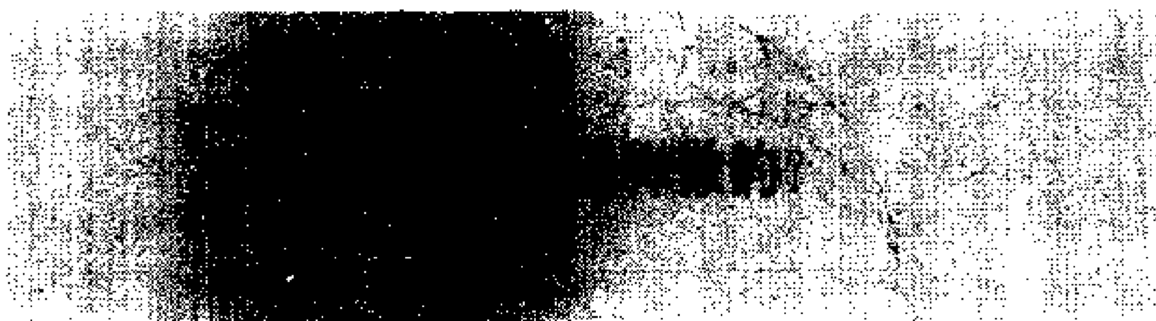
故此椭圆之面积为 $\frac{\pi ab(c^2-z^2)}{c^2}$,

表明椭球也可视为一种拟柱,于是我们可求得

$$V=\frac{4\pi abc}{3}.$$

就斯坦因梅茨立体而言,其体积是 $V=\frac{16r^3}{3}$. 有桩轶事提到电气天才查理·普鲁图斯·斯坦因梅茨(1865—1923),当人家问到这个立体(现在以他的姓氏命名了)的体积时,他根本不用任何纸笔,马上就信口说出正确答案,使在场的每个人都大吃一惊. 当人们问他何以竟能如此迅速地得出正确结果时,他却秘而不宣,守口如瓶. 现在人们相信,他大概是洞察到这一立体其实可以视为一种拟柱,而应用了拟柱求积公式之故.

原注:本文选自作者的系列讲演,题为《数学中的伟大时刻》,这是其中的一讲.



● 无任所数学家

[ ] 列昂·班可夫(Leon Bankoff)

解题是数学的一个主要乐趣,而解题的最大激励则在于获得一个漂亮结果,特别是通过奇妙手段来取得时. 巴布斯(Pappus)是一个优美定理的主人公,这点可由数学文献中无数次地提到"巴布斯定理"而得到证明. 毫无疑问,巴布斯成功地找到了一种巧妙办法来证明了他称之为"一条古典定理"的真实性. 这是一个他在其《数学汇编》第四卷里提到的命题,讨论内接于图形"鞋匠之刀"里的一系列相切圆. 这种图形首先被阿基米德在其《引理》中加以处理过. 我们可以满有把握地假定,这一切圆定理在以往只是作为经验知识来接受,但从未正式证明过.

在线段 AB 上任意选定一点 C,分别以 AC、CB、AB 为直径,在同侧作半圆,则三个半圆弧所围成的曲边三角形中的空间即称为"鞋匠之刀"形. 阿基米德研究过的"鞋匠之刀"的三性质之一是计算与三个给定圆弧相切的内切圆直径. 巴布斯从这个内切圆出发,大大地扩充了他的活动舞台,进而研究一系列相切的圆,其中的每个圆都与"鞋匠之刀"的两弧相切,如图 1 所示. 按照这个"古典定理",自第一个内切圆中心到直线 AB 的距离等于该圆半径的两倍;从第二内切圆中心到 AB 的距离等于该圆半径的四倍,……,对第 n 个圆,则是圆半径的 $2n$ 倍.

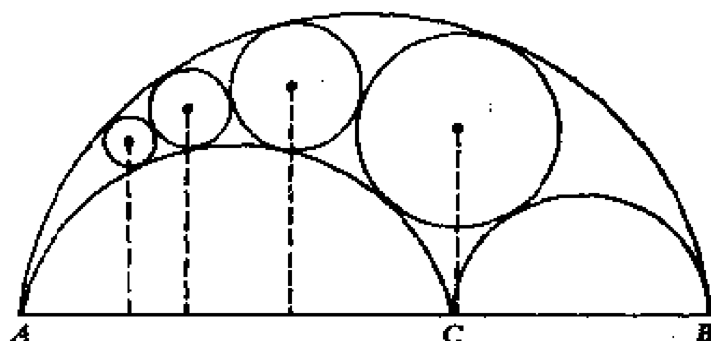


图 1

在图 2 的相关构形中，弧 CB 被删去，第一个内切圆切于弧 AC、弧 AB 及线段 CB. 再作出一系列相切的圆，它们都被混成三角形（其边可为曲线弧的三角形）所包围. 此时，从这一系列切圆的第 n 个圆的中心到基线 AB 的距离等于该圆半径的 $2n-1$ 倍. 例如，从第七圆中心到直线 AB 的距离等于此圆半径的十三倍.

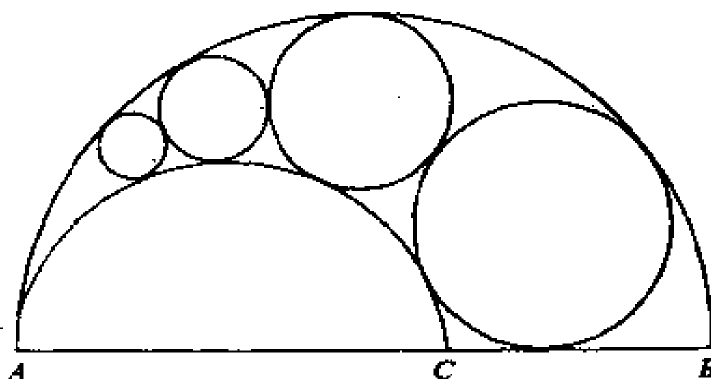


图 2

初看起来，这些性质的证明似乎十分简单. 我们这批二十世纪的数学家，由于手中掌握了复杂的工具而自鸣得意，把这类问题看作是按部就班的例行公事. 我们所要做的一切便是把反演原理应用于基本构形，然后让反演后的图形眼瞪瞪地望着我们，剥掉原来图形的神秘面纱，从而赤裸裸地显示出巴布斯圆定理的真实面貌.

为了照顾不太熟悉反演变换的读者,我们简要地介绍一下过程.自 C 点引 AB 的垂线,与半圆弧 AB 相交于 D.然后以 A 为圆心,AD 为半径,画出一段弧以表示反演圆的一部分.由于 $AD^2 = AC \cdot AB$,弧 AB 变成了在 C 点垂直于 AB 的 CD 的延长线.弧 AC 被反演为在 B 点切于弧 AB 的直线.由于反演圆与弧 CB 正交,所以半圆弧 CB 是自身反演的,它保持不变.对巴布斯的一系列圆来说,它们被变换成夹于两条平行的铅直线之间的一系列等圆,正如在原图中,被弧 AB 与 AC 所包围那样.在图 3 中,马上就可看出,自反演后的第 n 圆中心到基线的距离等于该圆半径的 $2n$ 倍.应用一个类似的反演变换于图 2,即产生结果 $y_n = 2n - 1$,这里 y_n 是位于基线之上的、第 n 圆中心的高度.

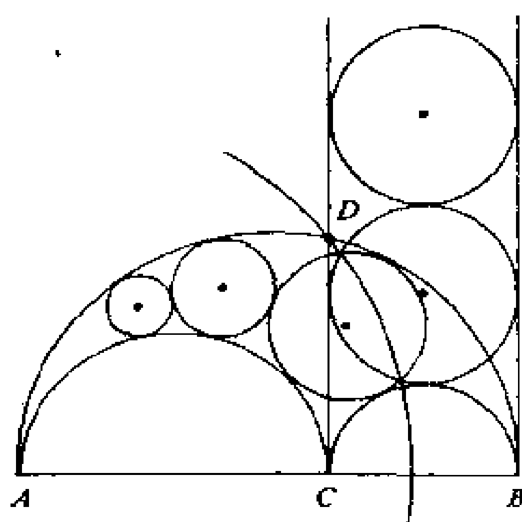


图 3

现在,让我们回到题目上的问题:“巴布斯是怎样做的?”他只能在没有反演(在他身后十五个世纪才出现)的帮助下来搞.为了取得他的惊人成果,欧几里得、阿基米德或阿波罗尼斯的著作中没有留下丝毫线索.因而巴布斯在着手攻打主要问题之前先着手证明几个有关的预备命题或引理.

我们表彰巴布斯的突出成就,也要向他道谢,因为他把复杂而

严密的证明遗赠给我们. 我们也赞赏他为了克服困难,解决问题而不得不锻造他自己的工具. 但是如果用现代的评估标准来看,我们将被迫承认,巴布斯所提供的证法并不是我们想提供给一个训练有素的数学家的方法,更不必说灌输给一位聪明而渴望求知的高中学生了. 如果我们让自发的好奇心追随巴布斯证法的思路,在其错综复杂的情况下吃力地前进时,这一点将变得越来越清楚. 但是我们从哪里去找这个证法呢?从巴布斯的日子到现在,十六个世纪已经过去,然而没有人愿意承担一件困难任务 —— 把他的著作《汇编》全部译成英语. 但是,巴布斯小部分著作的删节本已由托马斯·希思(Thomas Heath)勋爵译出,而在此书的标题中有着圆定理及其有关引理. 希思的学究式译文刊载于其著作《希腊数学史》,1921 年由牛津大学出版社刊行于世. 但是,在他的《希腊数学史缩本》(1931 年牛津版)一书中,他只是提了一下这个定理,巧妙地避开了问题解法的任何说明. 至于其余谈到希腊数学的著作,例如詹姆士·高尔(James Gow)与伊伏·托马斯(Ivor Thomas)的书,则干脆避免触及引理及主要证明.

我自夸私人藏书中有一册《汇编》的拉丁文本,它是 1659 年由康曼地诺(Commandino)从希腊文译来的. 为了对巴布斯的证明说上几句公道话,人们必须对有着十一页法定印张,密密麻麻地排满小字的论文来一个精耕细作,从其中搜索出巴布斯要说些什么. 另一方面,确实存在着一部两卷的法文本,该书由保罗·埃克(Paul Ver Eecke)忠实译出,精心编辑并由德·布路瓦与西(Desclée de Brouwer et Cie)印书馆于 1933 年出版于巴黎. 书中,十足有 19 页是专门描述巴布斯证法的. 在脚注中,保罗·埃克经常引证弗雷德列克·霍尔奇(Frederic Hultsch)的德译本,后者共有三卷,1876 年在柏林出版.

至此,读者已经得到授权,有资格了解巴布斯所用技巧的大要. 以下将通过四幅图(图 4 至图 7)进行简单介绍. 图 4 是说明引理一的,该图自身即可解释清楚 $KE \cdot EL = EB^2$,这一关系式下文将要用上. 这一结果的现代说法是:把点 L 与 K 取为关于 E 的反透射点.

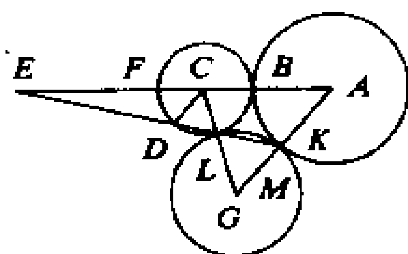


图 4

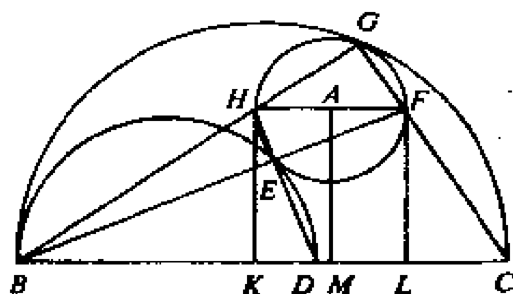


图 5

在图 5 中，圆心在 A 点的圆 FGH 是切于以 BD、BC 分别为直径（且 D 在 BC 上）的两个半圆弧的任意圆，作直径 HF 平行于 BC，其余作法看图就懂. 然后，由相似三角形 BGC，BKH 与 BLF，BED，建立比 $\frac{BC}{BG}$ 与 $\frac{BF}{BL}$. 巴布斯由此而得出 $\frac{2BM}{KL}=\frac{BC+BD}{BC-BD}$. 由于 $KL=2r$，这一关系式可写作 $\frac{BM}{r}=\frac{BC+BD}{BC-BD}$.

为了进入他证明的主体部分，巴布斯现在又建立起若干引理. 由相似三角形 BKH 与 FLC 他得出 $BK \cdot LC = AM^2$，又因 $\frac{BC}{BD}=\frac{BL}{BK}$，他又得到两个额外关系式 $BL \cdot CD = BC \cdot 2r$ 以及 $BK \cdot CD = BD \cdot 2r$.

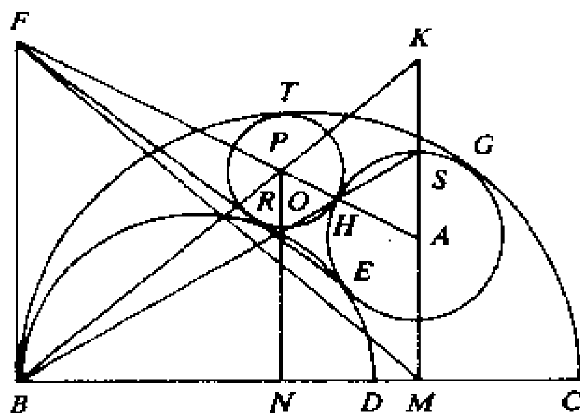


图 6

我们现在来看图 6. 如图，圆(A)与(P)是任意两个相切的圆，圆心 A 与 P 在 BC 上的射影分别为 M 与 N，巴布斯已阐明比$\frac{BM}{AS}$与$\frac{BN}{PO}$都是常数，且等于$\frac{BC+BD}{BC-BD}$，利用其他已建立起来的引理并在相似三角形中确认合适的关系式，他发现 $FH=FB$. 最后，通过更巧妙的手段，巴布斯成功地证明了 $\frac{AM+d}{d}=\frac{PN}{d'}$，这里的 d 与 d' 分别是圆(A)与圆(P)的直径.

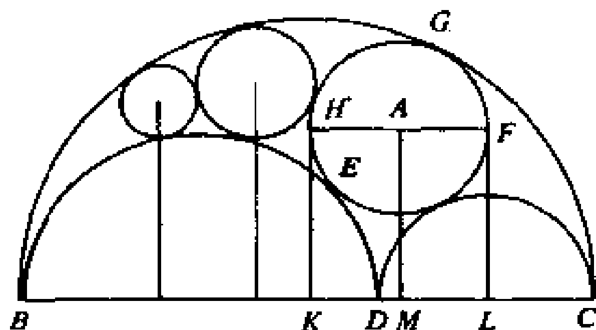


图 7

只是到了这一时刻，巴布斯才开始攻打主要定理. 参看图 7，并利用巴布斯巧妙地发现的引理，我们将能找到关系 $BK\cdot LC=KL^2$，$BK\cdot LC=AM^2$，因而 $KL=AM$，或 $p_1=d_1$，这里，p_1 是 A 到直线 BC 的距离. 设 p_2、d_2 表示第二圆的相应元素，则最后的引理将给出 $\frac{p_1+d_1}{d_1}=\frac{p_2}{d_2}$，于是得出 $p_2=2d_2$. 继续按同样方式进行，我们即可得到通过反演变换而得出的相同结论，即巴布斯系列中，任何一个圆的圆心到基线 BC 的距离等于其半径的 $2n$ 倍. 类似的论证也可以应用到图 2 的情形，结果表明，所求距离是相应半径的 $2n-1$ 倍.

以上是希恩所作的，说明巴布斯究竟怎样搞的简略提要. 它告诉我们，巴布斯怎样建立起骨骼，并在其上建立了他著名的圆定理. 如果我们以坚忍不拔的耐力全部吃透巴布斯的证明，我们就会确信巴

布斯真是一位最善于处理几何图形的天才.

犹如欧几里得与阿基米德,巴布斯纯粹采用几何方法来证明他的定理,只是偶尔一用最粗浅的代数,后者与比例式相较并不见得高明多少.我们已经说过,假如巴布斯懂得利用反演原理,他将如何证明其著名的圆定理.现在不妨让我们设想,藉助于我们现时的更灵巧便捷的代数手段,他将会怎样去做.

对图 8 采用新的记法.设 D、E 表示任意两个相切圆的中心(其半径分别为 x 与 y),它们夹在圆弧 AB 与 AC(其圆心为 O 与 P)之间.再设 F、H 表示 D、E 在 AB 上的投影.令 $AO=R$,$AP=r$,$DF=mx$,$EH=ny$.

由关系式 $FO^2-FP^2=DO^2-DP^2$,我们得出

$$(R-AF)^2-(r-AF)^2=(R-x)^2-(r+x)^2,$$

或

$$(R+r-2AF)(R-r)=(R+r)(R-r-2x),$$

于是有 $\frac{x}{AF}=\frac{R-r}{R+r}$.类似地 $\frac{y}{AH}=\frac{R-r}{R+r}$.

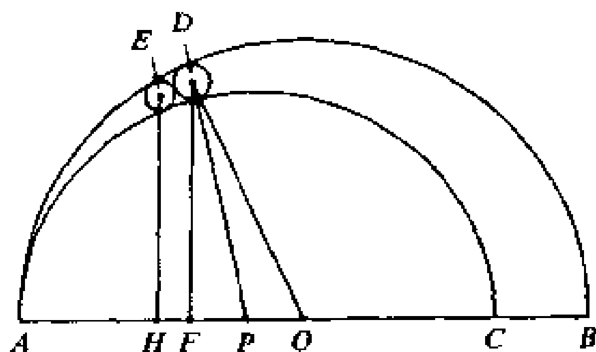


图 8

因而

$$HF=AF-AH=(x-y)\frac{R+r}{R-r}.$$

现在 $DE^2=HF^2+(DF-EH)^2$,所以

1. $$(x+y)^2=\frac{(R+r)^2(x-y)^2}{(R-r)^2}+(mx-ny)^2.$$

现在把三角形 DOP 的面积先用海伦公式表出，然后再用传统的"底高相乘积的一半"来表达，则可得到

$$\sqrt{Rrx(R-x-r)}=\frac{mx(R-r)}{2}.$$

类似地，对三角形 EHO，有

$$\sqrt{Rry(R-y-r)}=\frac{ny(R-r)}{2}.$$

求出 x 与 y，我们得到

2. $$x=\frac{4Rr(R-r)}{4Rr+m^2(R-r)^2}.$$

以及

3. $$y=\frac{4Rr(R-r)}{4Rr+n^2(R-r)^2}.$$

方程 2 与 3 给出如下关系

4. $$\frac{x}{y}=\frac{n^2(R-r)^2+4Rr}{m^2(R-r)^2+4Rr}.$$

现在，1 式可以化为下述形式

$$\left(\frac{x}{y}\right)[4Rr+m^2(R-r)^2]+\left(\frac{y}{x}\right)[4Rr+n^2(R-r)^2]$$
$$=2[2R^2+2r^2+mn(R-r)^2].$$

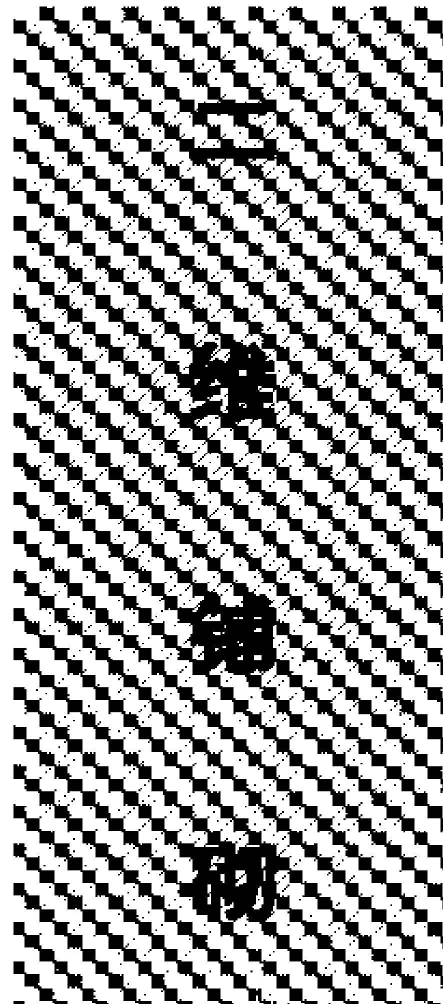
它与 4 式一起考虑，可以得出

$$4(R-r)^2=(R-r)^2(m^2+n^2-2mn).$$

最后，$m-n=\pm 2$，由于 $n>m$，所以我们得出 $n-m=2$.

请注意，D 既可选作初始圆的中心，又可选择为巴布斯系列中任意一个圆的圆心，而其比值 $\frac{EH}{y}$ 与相继切圆（一直扩展到无穷）的比值都得遵循模式 $n-m=2$，这就是说，相继比值属于一个公差为 2 的算术级数.

无论用几何、代数还是反演来求解，在过了许多世纪之后，巴布斯圆定理能够持续地激发我们的数学想象力，并以其独特的优美来引起我们的兴趣.


三维铺砌



● 贝尔实验室

□ R・L・格雷汉(R. L. Graham)

设想我们可以无限制供应大小为 2×1 的矩形砖块,并打算用它来铺 $p\times q$ 矩形房间的地板. 当然,我们必须完全铺满面积为 $p\times q$ 平方单位的地板,而且,任意两块砖都不得重叠. 例如,图 1 显示的便是 5×6 矩形地板的一种铺砖法.

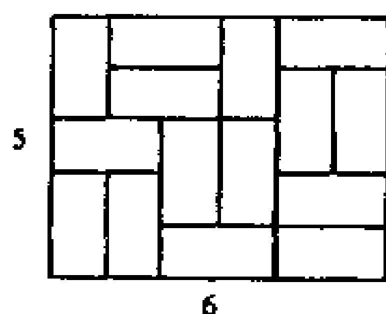


图 1

5×6 矩形地板的一种铺砖法.

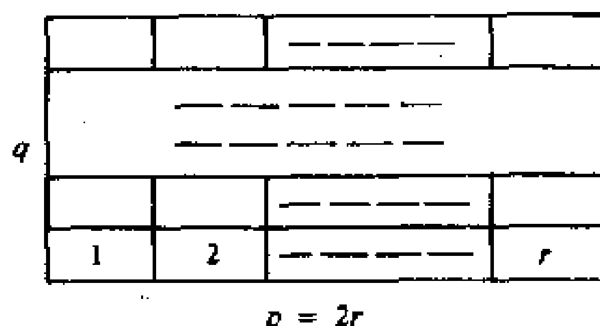


图 2

$2r\times q$ 矩形的铺砖法.

容易看出,如果这种砌砖法存在,则 pq 必为偶数,因为每块砖的面积等于 2. 另外,若 pq 是偶数,则 p、q 中至少有一个偶数,譬如说 $p=2r$. 在这种情况下,我们可以像图 2 那样铺砌.

无缝砌法

仔细察看一下图 1,我们就会注意到图中含有一条"裂缝",这就是说,有一条完全穿过矩形的直线,但却没有穿透任何砖块.让我们把不包含这种裂缝的砌法称为*无缝砌法*.如果我们把图 1 看作砖墙的横截面,那就很清楚为什么我们要尽量避免出现裂缝.在图 3 中我们给出了 5×6 矩形的一种无缝砌法.

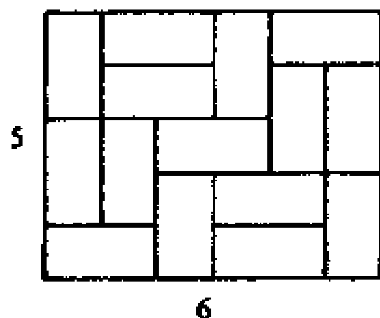


图 3

5×6 矩形的无缝砌法.

但是在人们试图造出 6×6 矩形的无缝砌法时出现了奇妙现象.(鼓励读者在往下阅读之前找一找它的解法.)对 4×6 矩形也出现了同样困难.事实上,对这两种情形都不存在无缝砌法!这就导致一个任何数学家都不能不问的问题[1]:

究竟哪一种 $p\times q$ 矩形存在无缝砌砖法?

回答问题

先从最明显的事实出发,如果一个矩形根本不存在铺砌方法(而

[1] 原注:另一个更一般的问题是:一个 $p\times q$ 矩形究竟有几种不同的无缝砌法?但此处我们不想讨论它.

积为奇数)，那它当然就谈不上什么裂缝不裂缝. 换言之，一个 $p\times q$ 矩形存在无缝砌法的必要条件是：pq 能为 2 整除.

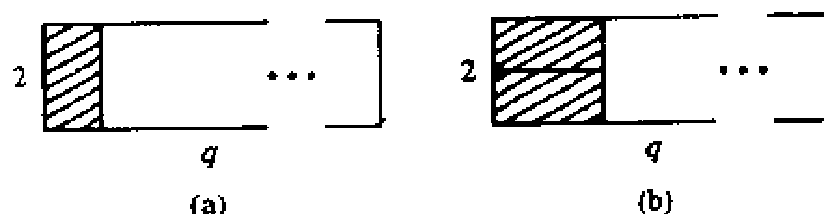


图 4

$2\times q$(?)矩形的无缝砌法.

然而，正如我们已看到的情况，这个条件是不够的. 例如，假定我们试图找出一个 $2\times q$ 矩形(此处 $q\geqslant 2$)的无缝铺砖法. 如图 4 所示，在该矩形的左边尽头，共有两种砌法. 但是，不论哪一种砌法(由于 $q\geqslant 2$)我们都将形成一条垂直的裂缝. 因此我们得出结论：即便其面积为偶数，这种矩形依旧不存在无缝砌法.

类似地，我们来考察在铺砌 $3\times q$ 矩形时，企图避免出现裂缝的尝试又将如何. 这时，从根本上说恰有两种办法来铺砌矩形的尽头(见图 5).

开始时如果照(a)那样铺砌将是根本不行的，因如 $q=2$ 就将出现一条水平裂缝，而当 $q>2$ 时又会出现一条垂直裂缝. 那么，如果照(b)的方式来铺砖，情况又将怎样呢？让我们在图 6 中显示其各种可

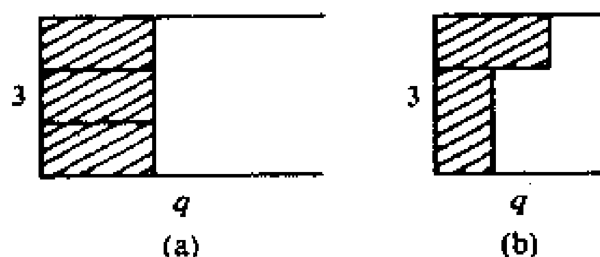


图 5

试图铺砌 $3\times q$ 矩形.

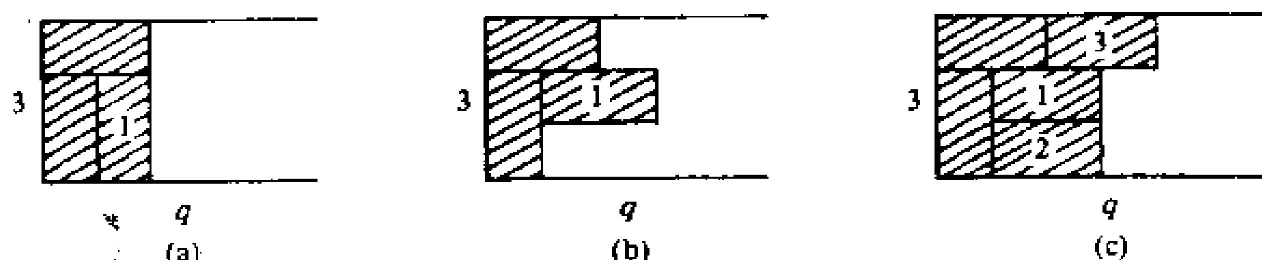


图 6

继续铺砌 $3\times q$ 矩形的尝试.

能性.

如前所述,情况(a)将会出现裂缝,必须舍弃. 剩下的可能性是把第 1 号砖像(b)那样放置. 可是,这将迫使把第 2、3 号砖照(c)那样放,否则势将无法填补(b)的放法所造成的缺口. 现在请大家注意图 6(c)模式的右侧轮廓,实际上它同图 5(b)的形状完全相同. 因此,我们仅不过是把图 5(b)所形成的缺口问题拖延了一下,最终我们仍然面对开始时的情况(毕竟,q 是有限数),而这个恼人的缺口只可能采用图 6(a)的办法,砌上一块垂直的砖来将其填满. 然而,这事一旦发生,我们马上造成了一条裂缝!

因此,我们的结论是:$3\times q$ 矩形不可能存在无缝铺砖法.

采用类似方式(但要多几种可能情况)可证明 $4\times q$ 矩形也不存在无缝砌法.(从对称观点可知,这意味着 $p\times 4$ 矩形也没有无缝砌法.)这时,人们也许抵挡不住这种诱惑,即把以上结论推广到 $5\times q$ 的矩形,等等. 然而,这是不对的,因为我们确实已找到了 5×6 矩形的无缝砌砖法.

到此地步,我们业已证明一个 $p\times q$ 矩形存在无缝砌法的另一必要条件是:p 与 q 都必须至少等于 5.

然而,这仍然解释不了何以 6×6 正方形不存在无缝砌法. 这个缺陷后来被果隆姆与杰威特(R. I. Jewett)的出色论证填平了.

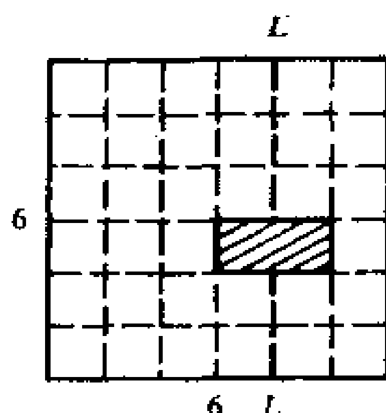


图 7

6×6 正方形中的潜在裂缝.

暂且假定我们已经找到设想中的 6×6 正方形的无缝砌砖法. 在正方形中存在着 5 条垂直与 5 条水平裂缝线，我们假定这些线都已被砖块打断(见图 7). 请注意每块砖正好打断一条潜在直线. 再进一步看，(这是至关紧要的一步!)如果任意一条裂缝(如像图 7 中的 L)正好被一块砖头打断，那么在此线两边的剩余区域，其面积必然是奇数，这是由于组成它们的 6×6 矩形都去掉了一个单位正方形之故. 然而，这类区域是不可能用 2×1 砖块铺砌的. 由此可见，每一条潜在的裂缝直线至少必须用二块砖头将其打断. 由于没有一块砖头能打断一条以上裂缝线，所以至少必须使用 20 块砖头. 然而 6×6 正方形的总面积也不过 36，可是 20 块砖头的面积却是 40! 于是，我们引出了矛盾. 所以 6×6 正方形的无缝砌砖法不可能存在.

到此地步，我们把已知事实总结如下：$p\times q$ 矩形存在无缝砌砖法的必要条件是：

1. pq 可以被 2 整除；
2. $p\geqslant 5, q\geqslant 5$；
3. $(p,q)\neq(6,6)$.

令人惊讶的是，这些条件同样也是充分条件. 这就是说，如果 p、q 满足条件 1,2,3，则 $p\times q$ 矩形必有一种无缝砌法. 要证明此理的一

种办法是从较小的无缝砌法起头，例如 5×8，6×8 或者我们早先说过的 5×6，从后再在它们的基础上构筑起较大的无缝砌法. 例如，假定我们已经有了一个 $p\times q$ 矩形的砌法，则图 8 告诉我们将怎样把它扩展成 $(p+4)\times(q+4)$ 矩形. 对任意正整数 r、s，一个与此相类似的构造法将能把原来的 $p\times q$ 矩形无缝砌法扩展为 $(p+2r+2)\times(q+s+2)$ 矩形.

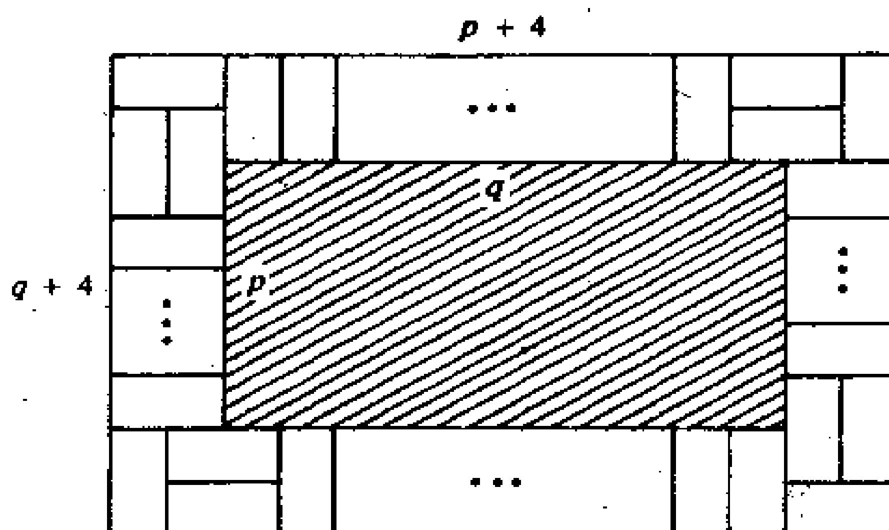


图 8

扩展无缝铺砌.

其他砖块

人们无疑会推测使用其他尺寸的砖块（如 3×1，或 7×5 的）时无缝铺砌的可能性问题. 初看起来，要明确指出怎么样的 $p\times q$ 矩形才能得到合适的铺砌，这似乎是一个毫无希望解决的困难问题. 然而，此问题却有着令人惊异的圆满回答. 如果我们采用的砖块大小是 $a\times b$，如有必要，我们可以改变量度单位，要使得 a 与 b 没有大于 1 的公因子.（我们当然也假定不去使用 $a=b=1$ 的那种砖，因为在那种情况下，将不再存在无缝砌法，除非那种只有 1 块砖的肤浅场合！）类似于前面所述的结果，它也存在两种基本的必要条件，其一是有关

可除性的条件，其二是有关大小的问题.

说到可除性，要铺砌一个矩形，不问其有无裂缝，都要求矩形的面积 pq 必须能被砖块面积 ab 整除. 其他要求也同样必要. 用 ab 种颜色对砖块与矩形以一种适当的轮回方式进行着色，可以证明下列条件是必要的：

1′. a 与 b 都必须至少整除 p、q 中的一个(这个条件要强于下述条件：ab 整除 pq).

至于大小问题，则一般情况同以前说过的条件 2(要求 $p,q\geq5$)也有个类比问题. 然而，新情况下的必要条件却是颇为出人意料的：

2′. p 与 q 的每一个都必须至少有两种方式表为 $xa+yb$ 的形式，此处 x、y 是正整数.

基本上，此点保证了在用砖块铺砌矩形时有足够的自由度，而不必老是只能用同样数目的砖块作水平或垂直配置.

条件 2′ 在 $a=2,b=1$ 的情况下就简化为上文的条件 2，这是由于 2,3,4 都不存在两种形如 $x\times2+y\times1,x、y>0$ 的表示法；然而，大于 5 的任一整数却是存在的，例如 $5=1\times2+3\times1=2\times2+1\times1$，等等.

奇妙的是，6×6 正方形不可能用 2×1 砖块无缝铺砌是唯一的特异例外. 在所有其他大小砖块的情况下，只要 p、q 满足必要条件，就可以利用上面已提到的技巧，通过类似步骤得到无缝砌法.

下面把一般结果综述一下：

定理 用 $a\times b$ 砖块铺砌 $p\times q$ 矩形的无缝砌法是存在的(我们假定 $pq>ab$，且 $(a,b)=1$)，当且仅当

1′. a、b 都必须整除 p 或 q 中的某一个；

2′. p、q 都能至少用两种方法表示为 $xa+yb,x、y>0$；

3′. 对 $\{a,b\}=\{1,2\},(p,q)\neq(6,6)$.

这一结果的详尽证法其实不难，我们把它留给热心的读者. 但是我们要交代一下，m 至少可用两种方法表示成 $xa+yb$，当且仅当 $m-ab$ 至少可用一种方法表示为 $xa+yb$(假定 $(a,b)=1$). 一般来说，这类整数

并不一定能像 $a=2, b=1$ 时那样可形成一个区间. 例如, 用 3×2 砖块可以对 11×18 或 14×15 矩形无缝铺砌, 然而对 12×12 矩形却不行($11=3\times 3+2\times 1=3\times 1+2\times 4$ 有两种表示法, 可是 $12=3\times 2+2\times 3$ 只有一种).

还要点什么?

在数学这一行里司空见惯的是: 一个结果往往引出 n 个更多的问题. 例如: 一个矩形究竟有多少种无缝砌法? 如果我们可用两种大小的砖块, 而不是只用一种, 情况又将如何? 在三维或更高维空间, 同样的问题将遇到什么情况? 如果每条裂缝都必须至少用两块砖打断, 情况将怎么样? 至少用 n 块砖打断呢? 在这个课题上, 我们已把读者带到现有知识的前沿. 我们鼓励有兴趣的读者自己去探索这一几何学的幽僻小径, 为他自己发现珍宝, 它一定静静地等待着发现它的"伯乐".



● 滑铁卢大学

□ W·T·塔特(W.T.Tutte)

有机会为这本祝贺马丁·加德纳的选集撰写文章，我感到十分荣幸. 从前，我曾在《科学美国人》杂志他的专栏上写过一篇专稿，内容是把矩形分解为若干个正方形的问题〔7〕. 或许再提供另一篇有关图形分割的文章会是合适的.

上面提到的那篇文章与"正方形集成的矩形"有关，也就是，要把矩形分解成若干个正方形. 一个矩形或正方形，如果被分割为许多个不相等的正方形时，则分别称作*完美矩形*或*完美正方形*. 那篇文章的根据是 Brooks，Smith，Stone 和 Tutte 等人的早期研究工作，曾于 1940 年发表在《杜克数学杂志》上〔2〕. 这篇早期论文给出了完美正方形的一个实例，并介绍了一种构作的方法(利用对称子图). 文中还有一个注记，提到可能的推广. 其中的一种推广是研究正三角形如何分解成其他正三角形的问题. 对此，我后来又写了两篇论文(〔6〕与〔8〕)，另一篇论文则是由 Brooks，Smith，Stone 与 Tutte 四人合写的〔3〕.

我没有任何理由去抱怨数学家与其他科学家竟会接纳那么多的、把矩形分解成正方形的论文. 他们对电流的基尔霍夫定律与纯粹几何的分割问题之间的神奇联系全都感到喜悦非凡. 但是，看来似乎对三角形分解问题就缺乏兴趣了. 然而，照我的意见，三角形分解问题其实是矩形分解为正方形问题的一种简单与巧妙的推广. 在本文

的其余部分我将试图证明这一点，并希望能唤醒公众对此问题的兴趣，而这种分解是理应受到重视的.

在我们讨论图形分割为正三角形时，原始图形究竟是三角形还是平行四边形(其内角为 60°与 120°)是无关紧要的. 一种三角形分割法即可变成平行四边形分割法，只要删去一个组成三角形(在一个顶点处)并加上另一个. 类似地，平行四边形分割法可以变成三角形分割法，只要在两条邻接的边上添加新的组成三角形.

平行四边形分割问题是矩形分割成正方形问题的自然类比. 设想我们从一个已经分解为正方形的矩形 R 开始，将它进行切变使之变为一个被分解为若干个菱形的平行四边形. 每个菱形可沿着一条适当对角线进一步分解为两个全等的正三角形. 这样一来，被分解为正方形的矩形 R 就转变成了一个被分解为正三角形的平行四边形 P. 正是基于此种观察，我们可以把矩形的正方形分解理论看作是平行四边形的正三角形分解理论的一个特例.

设 R 是一个完美矩形. 那么，把 P 称作完美平行四边形是否有意义呢?给出一个已分割为若干个正三角形的三角形或平行四边形，我们可以把它的一边看作“水平的”. 于是，任一组成三角形 T 就都具有一条水平边. 每个三角形可按其位置(在它的水平边之上侧或下侧)而规定它的正、负定向. 我们还应规定三角形 T 的“大小”，即它的每边长度. 其前的正、负号则取决于三角形的正负定向. 有了这些规定之后，我们就可以说，如果任意两个组成三角形的大小●都不相同，则这种三角形分割法就是“完美的”. 根据这一定义，我们就能说一个完美矩形 R 经过切变之后能得出一个完美平行四边形. 由 R 中的任意正方形变换出来的 P 中的两个全等三角形有着相反的定向，因而其大小视为不相同.

图 1 给出了一个不能从完全矩形经切变而得出的完美平行四边形. 在文献〔6〕中我很有点沾沾自喜，因为这个平行四边形具有最小

● 译者注：这里所谓的大小，不仅指绝对值，还应包括正、负号.

个数的组成三角形（一共十三个），又能满足完美性的要求. 为了同矩形的正方形分割问题进行对比，我喜欢把它的阶说成是$\frac{13}{2}$. 因为，矩形中的每个正方形都可以切割为两个三角形. 图 1 中，三角形中所标的数字表示的是它的大小.

在关于矩形的正方形分割的论文里，讲到这一步时，便用一张图来显示一个分划为正方形的矩形同电网络的关系. 下面的图 1 在平行四边形的正三角形剖分理论中也实现了同一目的. 不过，眼下，“电的”这一修饰词应按广义去理解. 此种形式的“电”在物理学中尚未发现（就我所知），不过我们认为它能够在有向图中流动，并服从基尔霍夫定律的某种类似法则.

平行四边形 P 画在图 1 的左边，相应的电网络画在右边. 网络 N 的每个顶点对应着一个极大水平线段，它来自 P 中组成三角形的

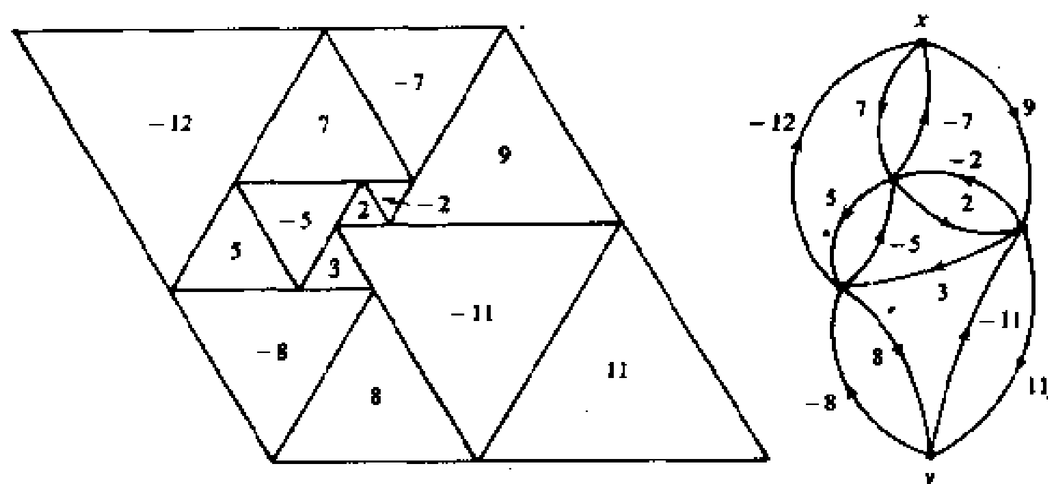


图 1

边长，在图中画于同一水平位置上. 网络是具有方向的，它的每一个箭头，也叫有向边，对应于 P 中的一个组成三角形（箭头方向从三角形的顶点指向底边）. 每个组成三角形的大小写在 N 中对应箭头的旁边，并称为这个箭头所负载的电流. 对应于 P 中上、下两条水平边

的 N 中的顶点被分别叫做 N 的**正极**与**负极**.

我们注意到 N 是一个平衡有向图("有向图"的英语单词为 Digraph,它是 directed graph 的缩合字). 平衡的意思是,在每个顶点,进入的箭头数与发出的箭头数相等. 不难看到,对任意 P 来说,N 是个平面图,它可以描绘在平面上,使得围绕每个顶点的箭头个数恰恰等于围绕在 P 中对应水平线段的三角形个数. 在这个图形中,围绕于每个顶点的入射箭头与发出箭头必须交替地出现. 我们称具有这一种性质的平面图为**交替图**.

在 N 的每一个顶点 v,所有指向 v 的箭头上所载电流之和称为 v **点的出 N 电流**. 其负值称为 v **点的进 N 电流**[1]. 于是,进入正极的电流应等于离开负极的电流,它们中的任一个都可以用来度量 P 的水平边的大小. 但是,在任意一个非极点的顶点处,进 N 电流与出 N 电流为零,而所谓的出电流正是 P 中对应水平线段上有其底边的三角形的边长的有向数字之和. 总之,在非极点的顶点处的电流法则就是我们对基尔霍夫第一定律的模拟规律.

让我们定义 P 的水平线段的电势为其高出于 P 的水平底边的距离,沿着 P 的一条斜边进行量度. 将此数转入 N,我们称之为对应顶点的**电势**. 与一个箭头 D 相联系的 N 的顶点分别称为 D 的头与尾,箭头的方向由尾指向头. 现在我们就可以把基尔霍夫第二定律的常见形式叙述如下:沿 N 的任一回路,不问箭头方向[2],总的电势降落为零. 我们注意到,一个箭头所载的电流在数值上等于从尾到头的电势差. 应用基尔霍夫第二定律我们就可以说,在任意闭合回路 C 中,电流之和为零. 当然我们事先需要约定,箭头中的流量,如与选定方向相反时,应取其负值.

---

[1] 译者注:读者当一字一句地用心阅读,并仔细领会此定义,以消除任何疑虑和误解. 特别应注意进 N 电流不是由发出去的箭头上所载之电流来定义的.

[2] 译者注:原文如此. 这句话说得不严密. 其实际意义是,只要能形成回路,可以先不问箭头方向. 但实际计算时,还是应考虑箭头方向的.

这里可以用上基尔霍夫定律的修正形式，正如其传统形式可用于矩形的正方形剖分一样. 画出一个交替图，在每个顶点至少有四个箭头，我们可以选择关联于同一个面的一对顶点作为正、负极点，然后可以从方程组解出一系列电流. 再像图1那样，回过头去求出相应的正三角形剖分. 如果碰上好运气，它将是完美的平行四边形. 这个过程的完整理论证明也许很困难，但其原理却十分清楚.

现在有可能从交替图推出正三角形分割的一些性质. 例如，我们可从欧拉多面体公式推出，对应于平行四边形的正三角形剖分的交替图要末包含一个2边的面，要末一个非极点的顶点只与4个箭头关联. 不论属于哪一种情形，平行四边形的组成三角形中，必有两个是合同的. 这一结论导致文献〔2〕中的严厉论断. 由于使用术语的差异，该文认为：*不存在完美正三角形.*

考察一个已作正方形分割的矩形 R 及其普通电网络 G. 假设它通过切变形成一个已作正三角形分割的平行四边形，并有了一个相应的网络 N. 我们发现：只要把 G 中的每条边代之以两条指向相反的箭头，就可从 G 得出 N. 在相应的交替图上，这两个箭头构成了一个二边形. 在作了单位电阻的假定之后，容易看出，在 N 上起作用的基尔霍夫定律的修正形式与在 G 上起作用的该定律的常见形式实际上完全等价. 因此，矩形的正方形分割仍然是三角形分解理论的一个特例.

先前在研究矩形的正方形分割中所得到的一些成果在目前或许会引导我们这样去想，有人大概已能造出一张清单，其中列举了一大批已作出正三角剖分的平行四边形了. 然而，实际情况并非如此.

这里没有篇幅可以让我们去详细讨论无向图 G 的基尔霍夫方程组. 我们注意到这一理论可建立在图的*基尔霍夫矩阵* $K(G)$的基础上. 假定 G 的各个顶点可列举如下：$v_1, v_2, \cdots, v_n$. 于是矩阵 $K(G)$具有 n 行、n 列. 它的第 j 个对角元是把 v_j 与其他顶点连结起来的边的个数. 而 i 行与 j 列的非对角元则是连结 v_i 与 v_j 的边数，再添上一个负号. 由此可见，$K(G)$是一个对称矩阵，每行与每列上各个元素之

和都是零.

矩阵-树图定理断言,如果我们从 $K(G)$ 中去掉第 j 行与第 j 列,则余下来的矩阵 $K_j(G)$ 的行列式是 G 的生成树数 $T(G)$. 让我们约定每条边的电阻都是 1 个单位,并从中选择两个顶点作为正、负极点. 把 $T(G)$ 看作在正极进入 G 或在负极离开 G 的电流大小是便利的,因为这样一来,在各条边上的电流就将都是整数. 这些电流对所选定的极点构成了 G 中的完备流动(满流). 这一满流中的电势差可表为 $K(G)$ 中某些子矩阵的行列式,再适当地乘上 +1 或 −1[1]. 作为其特例,极点之间的电势差是从 $K(G)$ 中去掉对应于极点的两行两列之后所得的子矩阵的行列式.

所有这些东西都可以令人满意地推广到有向图的不对称电学理论中去. 设 G 为任一有向图,不一定要求它平衡. $K(G)$ 的定义同上文一样. 第 j 个对角元是从其他顶点指向 v_j 的箭头数. i 行 j 列上的非对角元是从 v_i 指向 v_j 的箭头数,前面再添上一个负号. 我们须注意到这样的 $K(G)$ 不一定是对称矩阵,同一行的各元素之和为零,但对列来讲则不一定如此.

我们可以像以前一样地定义 $K_j(G)$,其行列式是 G 的生成树数,但并不是把 G 的所有生成树都计算,而只是那些每条边都从 v_j 引出去的生成树. 我们把这些结构称为 v_j 点的关于 G 的外向树枝图. G 的生成树,其中每条边都指向 v_j 的则称为 v_j 点的关于 G 的内向树枝图. (在无向图的场合,$K_j(G)$ 的值与 j 无关,但在有向图的场合,此情况未必为真.)

让我们选择一个正极 v_p,一个负极 v_q,并利用基尔霍夫定律的修正形式以计算相应的电流分布. 进入 v_p 的电流不一定等于离开 v_q 的电流. 尽管如此,我们仍然能找到一个满流,其中进入 v_p 的电流为 $\det K_q(G)$ 而离开 v_q 的电流是 $\det K_p(G)$. $K(G)$ 中一些适宜的子矩阵的行列式规定了这一满流中的电流与电势差. 人们发现这些电流与电

---

[1] 译者注:实际上就是代数余子式.

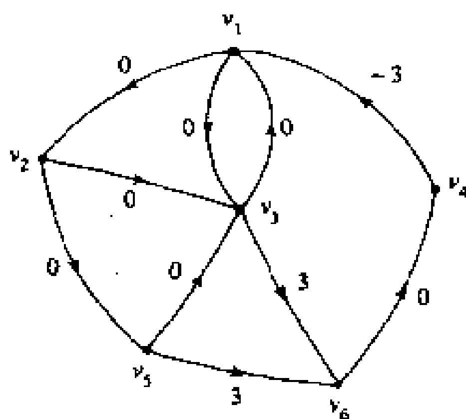


图 2

势差全是整数. 两极点之间电势差的规则与前面所说的一样.

作为一个实例,让我们来看图 2,把 v_1 选作正极,v_6 选作负极. 相应的满流也都已标明.

我们容易验证矩阵 $K(G)$ 应具有下列形状:

$$\begin{bmatrix} 2 & 0 & -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 3 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix}$$

让我们用 $T_j(G)$ 表示点 v_j 上的 G 的外向树枝图. 不论通过观察图 2,或者通过计算行列式 $\det K_1(G)$ 与 $\det K_6(G)$,都能发现 $T_1(G)=6$,$T_6(G)=3$. 因而图 2 中标明的流是满流,因进入 v_1 的流量为 3 而离开 v_6 的流量为 6. 从 v_1 至 v_2 的箭头以及 v_2 至 v_5 的箭头所载之流量为 0,这一事实是基尔霍夫第一定律修正形式的直接推论.

如果我们把每条边的指向都颠倒过来,以得出一个新的有向图 G',这时又会发生什么情况呢? 对一切非对角元来说,我们只要把

$K(G)$代之以它的转置矩阵. 但对角元有可能变动. 若 G 是图 2 的图，则 $K(G')$为下列矩阵：

$$\begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & -1 & 0 \\ -1 & 0 & 2 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}.$$

有向图 G' 由图 3 表示，仍然是有满流的.

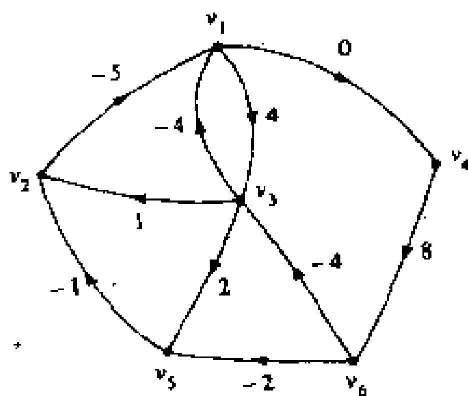


图 3

我们有 $T_1(G')=8, T_6(G')=9$. 图 G' 在 v_j 上的外向树枝图与图 G 在 v_j 上的内向树枝图之间显然存在着明显的一一对应关系.

当我们把理论局限于平衡有向图时，可以避开许多复杂情况. 由于 N 是一个平衡有向图，$K(N)$的元素，无论在每一行上或每一列上，其和均为零. 于是我们可以应用初等行列式理论证明，行列式 $\det K_j(N)$与 j 无关. N 上一个给定顶点 v 的外向树枝图个数，对每个 v 都是相同的. 于是我们提到这个数时即可简单地称为 N 的树数 $T(N)$. 我们可以把图 1 的平衡有向图取作一个实例. 如果图上各顶点按照自上至下的顺序来编号，则有

$$K(N)=\begin{bmatrix} 2 & -1 & 0 & -1 & 0 \\ -1 & 3 & -1 & -1 & 0 \\ -1 & -1 & 3 & 0 & -1 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}.$$

由此可推知，在满流时，极点之间的电势降落为

$$\begin{vmatrix} 3 & -1 & -1 \\ -1 & 3 & 0 \\ -1 & -1 & 3 \end{vmatrix}=20.$$

于是我们推论出图 1 中的流是满流. 从而 $T(N)$ 等于进入正极的电流，所以它等于 19.

给出一个平衡有向图 N 之后，我们可以把它的所有箭头完全颠倒，从而得到一个有向图 N'. 显然，N' 也是平衡图. 另外，如果 N 定义平面上的一个交替图，则 N' 亦然. 但既然 $K(N')$ 仅不过是 $K(N)$ 的转置，于是 $T(N')=T(N)$. 因此，对 N 的每个顶点来说，内向树枝图与外向树枝图的个数相等.

设有两个顶点在 N 与 N' 中都被选作正负极点，现在考察两个满流. 由于树数相等，所以在两个图中，进入正极的电流都一样. 另外，两个极点间的电势降落也一样，因为，在 N 中，它是由对称于主对角线的子矩阵的行列式所给出，而在 N' 中则由该矩阵的转置矩阵的行列式所给出.

如果 N 对应于一个有正三角形剖分的平行四边形，则 N' 亦然. 由上节的结果，可知平行四边形 P 与 P' 应具有同样的大小和形状. 不幸这种效应不能通过图 1 的有向图来阐明. 把它的箭头全部颠倒之后，仅能得到一个与原图同构的有向图. 而且每个极点在此同构下不变. 若 N 对应于一个具有正方形剖分的矩形，则 P 与 P' 对应于 R 的两种切法. 在切变时，我们既可以把上面的水平边向左移动（相对于下面的水平边而言），也可以向右移动.

在文献〔6〕的 478 页上，画出了对应于一个交替图 N 的两个平

行四边形 P 与 P'，它们都已作了完全的正三角形剖分. 每个平行四边形的水平边长为 3441，倾斜边长 2999. 两者都被分割为 36 个组成三角形，其大小均不相同. 两者都有两个合同的组成三角形，其边长为 129. 但除此之外，再也没有其他一样尺寸的正三角形了.

与平行四边形的三角形剖分理论有联系的不对称或有纰漏的电气性质在文献[3][6][9]中有更详尽的讨论. 在大多数一般理论中，一个箭头的电导（电阻的倒数）不限于 1，而可以是介于整数之间的中间值.

具有正方形分割的矩形有两条互相垂直的边，其中任一方向都可视为水平. 两种不同选法将会导致两个电网络 G 与 G_1. 它们之间有着简单的联系. 每个图都可通过添加连接极点的一条新边而变为完全. 这两个完全网络（通常称作 c 网络）是平面对偶图形. 由于两个相应的电网络方程组会得出同样的正方形分割模式，这就使人们有理由去猜想行列式 $\det K_j(G)$ 与 $\det K_j(G_1)$ 是相等的. 现在已经可以明确肯定这个想法是成立的. 可以证明互为对偶的平面连通图具有相等的树数.

在三角形分割理论中有没有与一对对偶 c 网络相对应的东西？回答是的确有的，但为了显示它，我们必须从三角剖分的平行四边形过渡到三角剖分的正三角形. 作为一个例子，我们可以在图 1 的三角剖分平行四边形P的基础上添加两个新的组成三角形，于是就得出

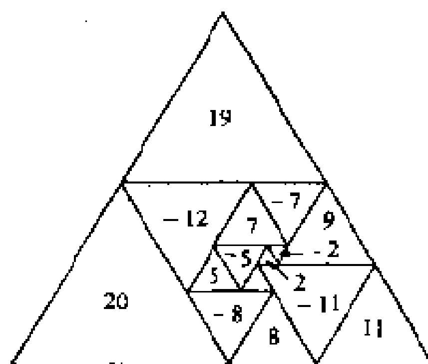


图 4

了图 4 中的有三角剖分的正三角形.

图 1 那个电网络的相应修正是明白易懂的. 为了表示大小等于 20 的新的正三角形,我们要在左边新添一个箭头,其指向由原来的正极 X 到原来的负极 Y. 相应于大小为 19 的三角形,我们要有一个新顶点 Z 以及自 Z 指向 X 的新箭头. 我们于是可以说 Z 表示有三角剖分的正三角形之顶点,即使这个顶点并不是严格意义下的水平线段.

新的电网络中的电流将会服从基尔霍夫定律的修正形式,此时 Z 作为正极而 Y 作为负极. 从 Z 引出的外向树枝图数可以算出为 39,它是分割的三角形之边长. 从 Y(或者除 Z 以外的任何顶点)引出的外向树枝图数显然为零. 这对应于下列事实:在 Z 点进入网络的电流是零.

在电网络图上使 Z 点与 Y 点重合将带来一些方便,这样一来图形会显示出一种新的交替图 M. 原先与 Z 关联的箭头可以称为极箭,在图上将用一根打上交叉记号的横杠来标明. 图 5 就是与图 4 的三角形剖分相对应的这类电网络图. 其中各个顶点与三角形剖分中的水平线段相对应.

图4与图5显示了一般过程的一个实例. 给出任意交替图 M 之后,我们可以在其上标出一个极箭 D,这时就可以指望从它推导出一个相应的具有三角形剖分的正三角形. 为了形成对应的电网络,我们要在D中分离出它的尾顶点Y,并重新给它一个与其他箭头都无关

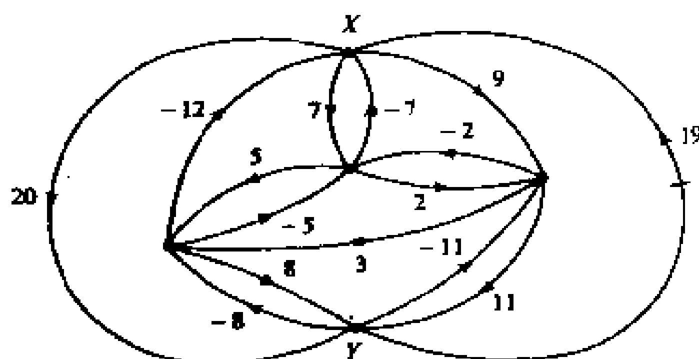


图 5

联的、新的尾顶点 Z. 这样做了之后，再把 Z 取作新的正极，Y 取作负极. 由此算得之电流值即可解释为各个组成三角形的大小.（有时我们会算出一个值为零的电流，这时就将引起疑难情况.）

回到图 4，我们注意到它的组成三角形的各边形成了三族平行线段，其中任一族都可选作水平线. 譬如说，我们可以考虑图 4 中平行于左边那条斜边的线段，并使这些线段与另一个电网络中的顶点相对应. 这样的图形已在图 6 中表示出来，还有一个已打上标记的极边. 至于对应于右面一条倾斜边的一族线段，则可产生图 7 那样的电网络.

我们找到了三个不同的电网络图 5、图 6 和图 7，它们都载有同一组电流. 交替图的这种三合一现象可看作矩形的正方形分割理论

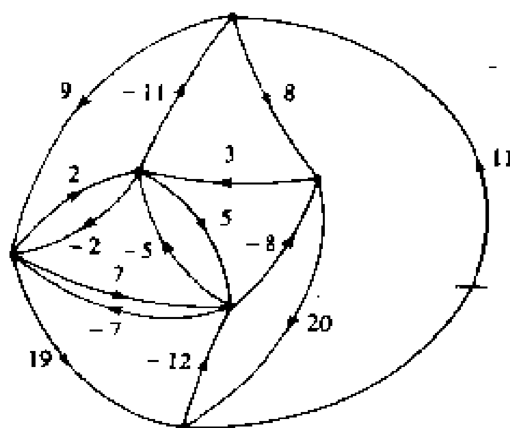


图 6

中对偶图的类似物. 在文献〔6〕使用了“triality”这个单词，文献〔8〕中则使用了“trinity”这个单词. 产生三合一现象的三个图在文献〔8〕中称作“互为 trine”❶. 给定任意交替图之后，我们可以直接定义它的三个“三合一”图，而不必绘出一个与之关联的三角形剖分图. 在文献〔8〕中，阐明了怎样从一个双三次图 M 中推导出交替图 M_1，M_2，M_3

❶ 译者注：这几个英文单词都具有“三位一体”或“三合一”的意思. 都不是常用词而且词形各异，相当于我国的异体汉字. 作者在此大讲特讲，对我国读者则可不必深究.

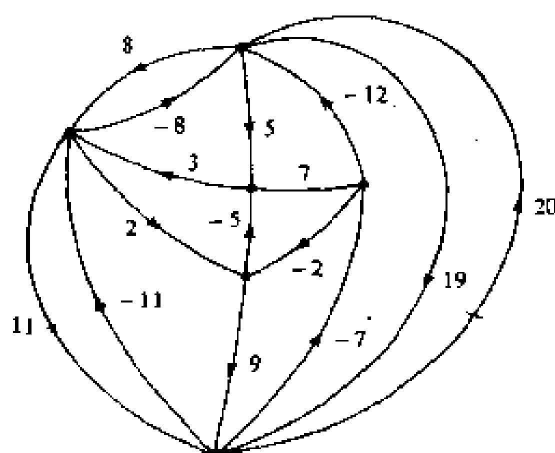


图 7

来，并讲明了从三合一图（联系一个特定的三角形剖分图）又怎样找出对应的双三次图 M.

在一个双三次图中，顶点被区分为黑、白两类，每条边把一个白顶点与一个黑顶点连结起来. 此外，面可以有 3 色（例如红、绿、蓝），从而每条边可以分开两个不同色的面. 除了三色的各种不同排列之外，3 色着色法是唯一的. 我们可按如下方法作出相应的三位一体图的交替图 M_1. 在 M 的每个红色面中给它一个顶点. 再从 M 的隔开绿色面与蓝色面的边中形成它的箭头. 每个箭头的方向则从黑的一个端点指向白的一个端点. 然后，把每个箭头都加以延长，从穿越 M 的邻接红色面的一端延长到该面中所包含的 M_1 的顶点. 三合一图的其他成员也可作类似之定义，每一个都是这三种面色的一种重新排列.（在文献[8]中也给出了三合一图的定义，但不是直接利用双三次图 M，而是利用了它的对偶图. 这个对偶图是球的一个三角剖分.）

设 M_1 是从无向平面图 N 中把每条边代之以两条相反的箭头而得出之图. 后来发现三合一图的另一个成员（如 M_2）也可以用类似办法从 N 的对偶图中导得. 三合一图的第三个成员是 N 的中间图的一种有向形式. 这一观察证实了我们的论断：三合一性乃是对偶性的

一种真正推广.

由于对偶图具有相等的树数，我们可以指望三重交替图也会有着相等的树数，这一想法确实已被证明为正确.（在文献[6]与[8]中给出了证明. 一个较简证明最近由 K. A. Berman 发现，见文献[1].）

三个三合一交替图 M_1, M_2, M_3 所共有的树数可以认为是有关双三次图 M 的一个性质. 在文献[3]中由 M 的结构直接给出了它的解释，作为匹配的计数，而不是作为树的计数.

在这里结束我们的补充说明看来是时候了. 对进一步研究感兴趣的读者可以查阅我们已列出的参考文献. 但有一个细节需要特别注意. 在三角化分割中可能出现六个组成三角形交于一点的情形. 这一点称为一个叉点，它对应于矩形的正方形分割中的叉点. 其时，四个组成正方形交于该点. 无论是哪个理论，通常都不大可能在一个有趣的实例中遭遇一个叉点，除非在作图中强行作出某些对称处置. 但是，在叉点确实出现时，那就有必要引入新的规定以保证仍能有成对的对偶电网络或者三合一的三个交替图. 要而言之，我们要对叉点处的一些最大线段进行划分. 只有划分后的子线段才能同电网络的顶点一一对应. 在三角形的场合，每个叉点是三条最大水平线段的交点. 其规则是，其中的两条线段（可任意选择）需要继续划分.

补　　遗

写作本文重新复活了我对平行四边形三角剖分的兴趣，于是我马上动手再算它一、二个. 结果我算了两例，看来它们都值得记录下来. 它们都是完美的三角剖分，其阶均为 $11\frac{1}{2}$. 每个平行四边形的水平边长为 401，斜边长为 264. 同等大小的组成三角形，在一个图形中出现时，另一个图形中就不会出现. 对我来说，同一平行四边形有着两种不同的完美三角化分解，又符合以上情况的，只知道这个唯一的例子.

两者的电网络图具有紧密联系，只要把箭头全部颠倒一下，即可从一图得出另一图. 这些网络与它们的流量都已在图 8 与图 9 中给出. 我已校验过，图中之流是满流. 在此两图中，X 是正极，Y 是负极. 进入 X 与离开 Y 的流量为 401，从 X 到 Y 的电势降落为 264. 我想，现在，读者将能毫无困难地从电网络图画出实际的三角形剖分图.

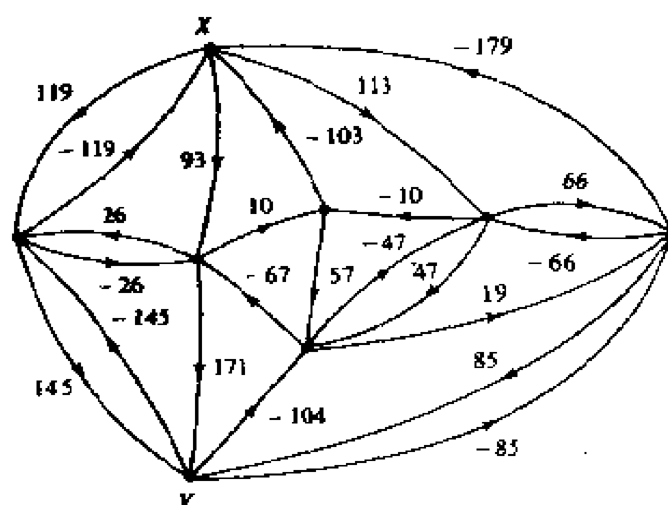


图 8

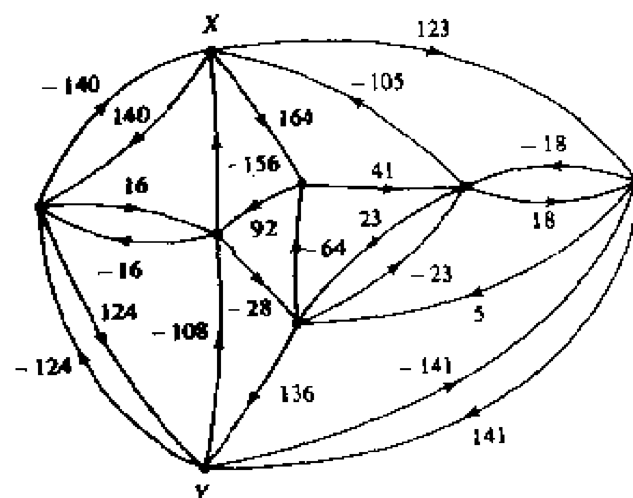


图 9

参考文献

1 Berman, K. A. 1978. Spanning trees, arborescences and 4-valent graphs. *Thesis*. Waterloo.

2 Brooks, R. L. ; Smith, C. A. B. ; Stone, A. H. and Tutte, W. T. 1940. The dissection of rectangles into squares. *Duke Math.*, 7 : 312—340.

3 ________. 1975. Leaky electricity and triangulated triangles. *Philips Res. Reports* 30 : 205—219.

4 Duijvestijn, A. J. W. 1978. Simple perfect squared square of lowest order. *J. Combinatorial Theory B* 25 : 240—243.

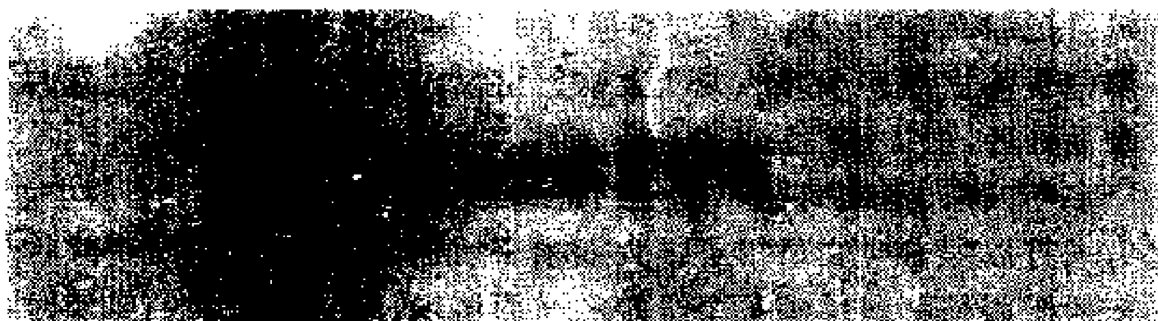
5 Sprague, R. 1939. Beispiel einer Zerlegung des Quadrats in lauter verschiedene Quadrate. *Math. Zeitschrift* 45 : 607.

6 Tutte, W. T. 1948. The dissection of equilateral triangles into equilateral triangles. *Proc. Cambridge Phil. Soc.*, 44 : 463—482.

7 ________. 1961. Squaring the square. In *The 2nd Scientific American Book of Mathematical Puzzles and Diversions*, ed. Martin Gardner. New York: Simon and Schuster.

8 ________. 1973. Duality and trinity. *Colloquia Mathematica Societatis Janos Bolyai*, 10 : 1459—1472.

9 ________. 1976. The rotor effect with generalized electrical flows. *Ars Combinatoria* 1 : 3—31.



莫拉维安学院

□ 多丽丝·沙特斯奈德(Doris Schattschneider)

马丁·加德纳的专栏"数学游戏"的最吸引人的方面之一是它介绍了一些数学问题,这些问题是为了引起业余爱好者们的兴趣以及鼓励他们的解题努力而专门设计出来的. 由于他自己的一贯坚持,加德纳自己也是一位业余数学家,而且又对正规的数学教育并未表示多大敬意——在他的专栏文章里,对数学里的巨头和无名小卒的业绩都是一视同仁地歌颂,他们的名字也往往并驾齐驱,没有加上头衔以资区别. 业余爱好者们是他最起劲的追随者,他们勇敢地接受一些题目的挑战,并由于发现了足以与其智慧相称的、独特的解题方法而自得其乐. 令人惊讶的是,他们虽然缺乏正规的数学教育,但这往往是其长处而非短处,在解决问题中所表现出来的巧妙解法有时竟会凌驾于专业人员之上.

证明上述论点的一个突出例子是加德纳在 1975 年 7 月所写的专栏文章"用凸多边形砖块铺满平面"所引起的后果. 该文激起理查德·詹姆士(Richard James)的强烈兴趣,他决定在解法上露一手. 其结果是,他的浅显的方法(与人们熟悉的方法不一样)产生了一种正规数学体系所忽略的解. 其后詹姆士作出新发现的报道又在玛乔莉·赖斯(Marjorie Rice)的身上引起了巨大兴趣,她以坚韧不拔的耐力对之进行了透彻研究(绝大部分是在她家厨房里的锅台上作出的),

结果终于导致一系列丰硕新成果的出现. 约请撰写本文的邀请使我有机会把加德纳的一篇专栏文章所造成的事件细节联系起来——这些事件目前依然在平静的湖面上吹起涟漪. 这要归功于成千上万的业余爱好者, 正是他们使加德纳的专栏文章取得了这样的成功.

多年以来, 铺砌问题一直是加德纳喜爱的课题. 最近二十年中, 已有一打以上专栏文章用了很大篇幅来讨论它. 怎样用一些薄片状的东西服服帖帖地铺满某个给定的平面区域, 看来这是我们童年时代的消遣之一. 即使已经成人, 我们也会由于纯粹娱乐或出于实际需要, 继续被这些砖、瓦之类的铺砌问题所吸引. 一个最基本的铺砌问题是:"究竟是何种形状的瓦片, 通过不断重复使用, 可以铺满平面, 使之既无空隙, 又不重叠?"于是许多明显的形状便会涌上心头——不需要什么数学家来为你提供一串长长的清单. 在世界各地古代镶嵌工艺品中发现的一些形状美丽的瓦片充分表露了装饰艺术家在解决该问题中的想象能力. 对我们的问题, 最一般的答案迄今还是未知的. 为了获得部分解答, 需要对瓦片的形状作出一些规定, 而这些问题的答案也在寻觅之中. 例如我们有如下一些问题: 如果瓦片是由正方形组合起来的(多连骨牌), 能否用它们铺满平面? 如果由正三角形组合起来(复合三角块), 情况又将如何? 由正六边形组合起来的形状(多连六边形)呢?

加德纳 1975 年 7 月的文章讨论了铺块是凸多边形的问题. "什么样的凸多边形能铺满平面? 要讲清楚加之于凸多边形的条件以保证做到这一点."这样, 加德纳选中了一个已被许多数学家研究了五十多年并被宣布已经彻底解决了的问题. 容易看出任何三角形或任何平行四边形都能铺满平面, 但是凸五边形或边数更多的凸多边形却并不是都能做得到. 例如, 正五边形不可能铺满平面, 可是任何五边形, 只要有一对平行边就能办得到. 正六边形可以铺满, 但别的一些六边形却不行. 七边形或边数更多的凸多边形也不能铺满. 最后的这一论断曾被加德纳宣布为"不难证明", 并被人家说成是数学上的一个"民间的定理". 这样一来, 每个人都在引证它, 但似乎没有一个

人能给出它的一个完整而确切的证明. 幸而,伊凡・尼文(Ivan Niven)的最近一篇文章发表于1978年12月号的《美国数学月刊》上,像是已经填补了数学文献方面的一个缺陷,提供了一个透彻而深具说服力的证明.

加德纳的文章奠基于R・B・克希纳(R. B. Kershner)1968年的论文,该文对问题"什么样的凸多边形可以铺满平面"的一切答案作了全面检阅. 六边形情况下的完整答案(共3种类型),五边形情况下的一部分答案(共5种类型)已被K・来因哈脱(K. Reinhardt)于1918年先后发现. 克希纳的广泛研究又得出了三种类型的五边形,它们也都能够铺满. 克希纳感到这张表格目前已经完备(见图1). 他对于该问题的强烈爱好见于一封曾被加德纳引证过的信:"由于很难解释的某些原因,我被这个问题困扰已经长达三十五年之久. 每过五年、十年,我就要作某种尝试,企图一举解决问题. 大约两年以前,我终于发现了一种分析五边形铺砌可能性的方便办法,它要比来因哈脱的更为优越. ……这项研究的结果是:发现了另外三种类型的五边形,它们同样能够铺满平面……. 这些铺砌方法委实令人惊讶之至. 发现它们的存在令人深感喜悦. "克希纳对本问题的迷恋后来被证明为具有感染性. 加德纳的讲解把该问题从尘封的数学杂志里(多年来已经无人问津)"请"了出来,把它置于广大读者之手,其中也包括许多趣题爱好者. 我们的故事就打从这里开始.

当理查德・詹姆士第Ⅲ看到加德纳的文章时,他只读了第一部分. 他决定在继续阅完文章之前,先测试一下他自己的解题技能. 他在信中(考克塞特(H. S. M. Coxeter)善意向笔者提供了该信件)写道:"在读完加德纳先生的、记述克希纳先生八种铺砌五边形的文章之前,我决定自己动手试一下. 涌上心头的第一桩事是……取一些正八边形(用正方形填满隙缝),然后再予以调整以便改用五边形取代正方形(需将正八边形移出"格子",转为"平行"的带条形状). 正八边形正好可以天衣无缝地分割成四个五边形. 划分法可以记录如下:

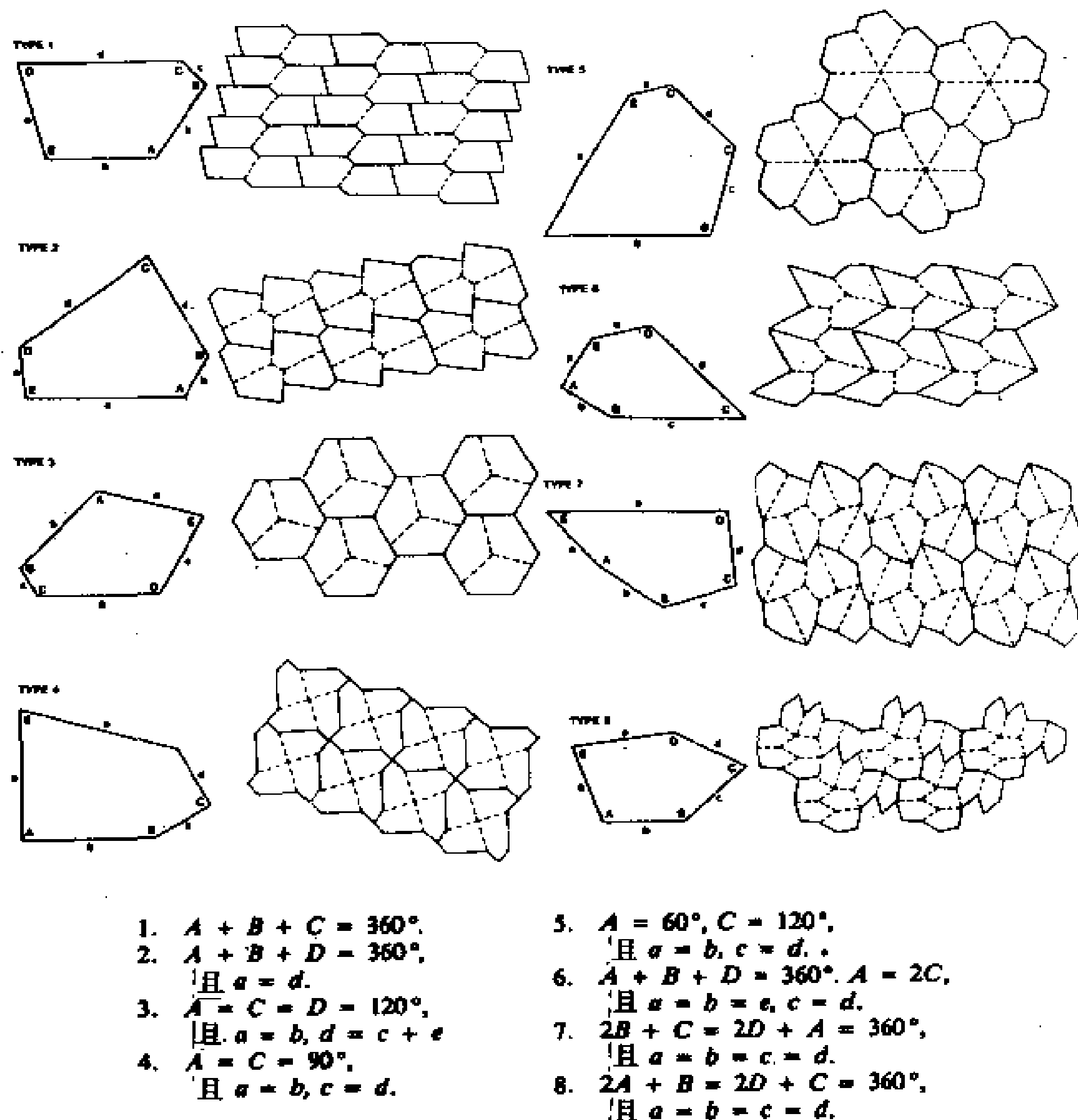


图 1

马丁·加德纳在 1975 年 7 月所报道的、能够铺砌平面的、八种类型的五边形.

$A=B=E=90°, C=D=135°, a=b=2c=2e$. 此种铺砌法太有趣了，不过所用的五边形却是很不起眼的. 旋转正八边形中出现的交叉十字（并作其他的必要调整），结果就产生了（五边形以及图 2 的铺砌法）.”詹姆士把他的发现送给马丁·加德纳并向他发问：“你是否同

意克希纳先生遗漏了这种铺砌法?"令人兴奋的消息立即由加德纳通报给克希纳与其他几位数学家. 克希纳的幽默反响以及詹姆士的拼镶法在1975年12月的《数学游戏》专栏上发表出来,告诉了读者. 一位业余爱好者通过一个新发现的例子证明铺砌五边形的那张清单尚不完全. 是否还存在其他可能性呢?

也许在这里可以稍为岔开一下,对加德纳传播的"数学小道新闻"(我将简单地记之为MG^2,即马丁·加德纳的数学小道新闻,请参看图3[1])略为瞟上一眼. MG^2的嗡嗡声几乎维持了一年,经常有大批信件蜂拥而至,还有电话呼唤与私人交谈. 在研究或撰写专栏文章时,加德纳经常与专家学者或掌握这个课题的最新信息者频频接触. 发表文章之前,对其初稿的正确与否先征求他们的意见. 反之,当加德纳收到来信,在把它存档备查之前,他先要把复印件送给他认为对之深感兴趣的人. 通过此种方式,最新消息提供给了那些关心者,发现的正确性经过核查,加上评论后退回给加德纳,更重要的是,在同一问题上的研究者通过加德纳而彼此有了联系. H·S·M·考克塞特是从加德纳那里获得詹姆士的新发现的一份复印件的,正是从考克塞特那里我得知该发现的信息. 我已无法拒绝在这个问题上试一试自己的手艺——不久我就获得可以铺满平面的一族五边形所应满足之条件,这就推广了詹姆士的结果,他送给加德纳的不过是该族五边形中的一个个别例子而已.

按照克希纳的记录法,詹姆士型五边形的一般形式应满足下列条件:

$$A = 90^\circ,\quad E = 180^\circ - B,\quad D = 90^\circ + \frac{B}{2},$$

$$C = 180^\circ - \frac{B}{2},\quad a = b = c + e.$$

[1] 译者注:原文为 Martin Gardner's mathematical grapevine,其词头缩写为 MGmg,即MG^2.

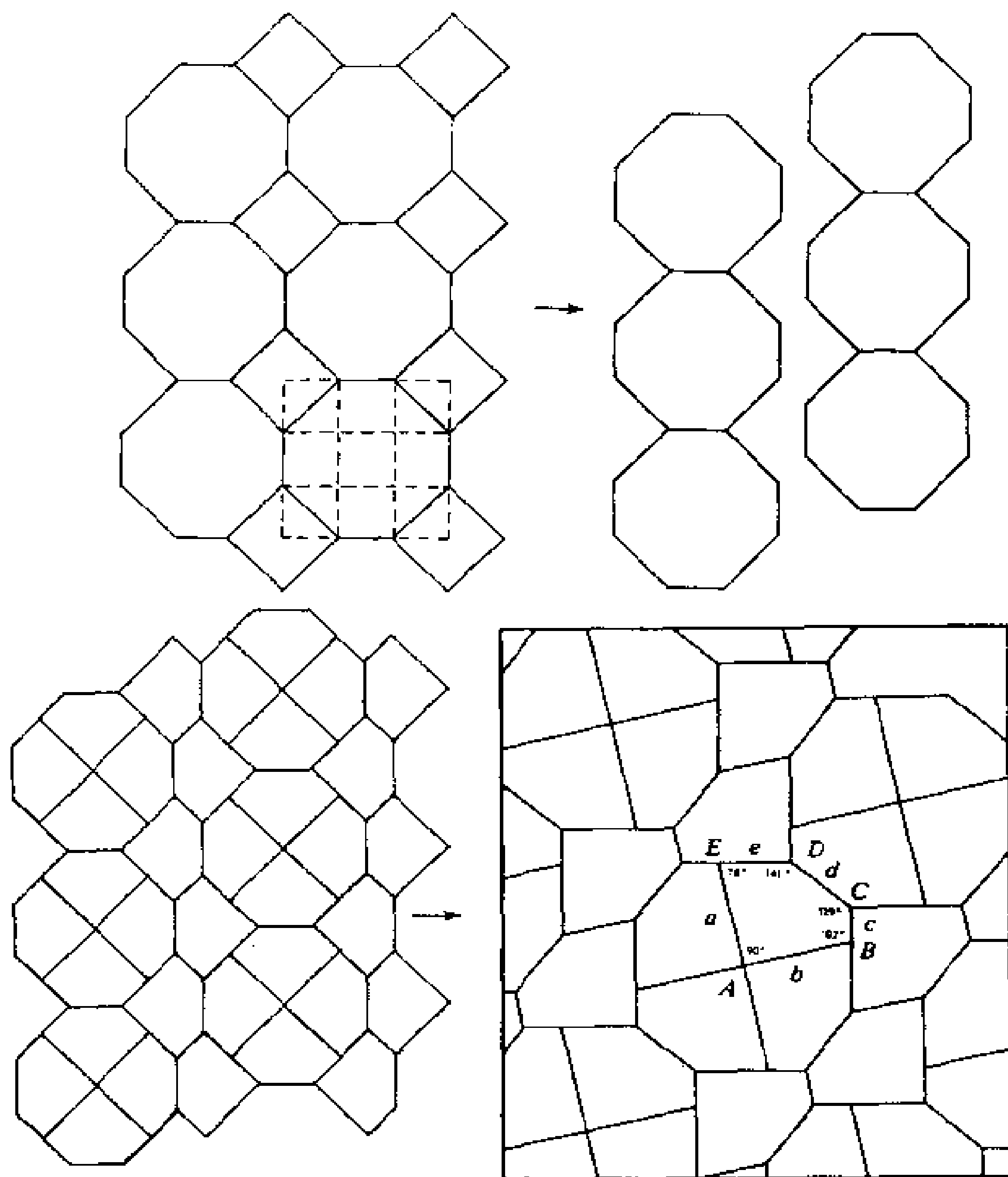


图 2

理查德·詹姆士怎样发现一种可铺满平面的新的五边形.

从常见的正八边形与正方形(用虚线给出其作法)铺砌法出发.先进行平移,观察一下五边形能否取代正方形的位置.一种成功的新铺砌法于是产生出来.进一步把正八边形分割为五边形,最终得出了一个以前没有发现过的、可用作基本铺砌材料的五边形族.

当我把这些东西送给加德纳与考克塞特之后，我发现自己已变成加德纳“五边形问题”往来信件的一位收件人，并与这个问题上的活跃研究家们取得了直接联系．如果我在此问题上的日益膨大的文件档案可以看作加德纳的一篇专栏文章所引起的读者来信的指示器，那么他的档案肯定会装满几间房子！



图　3

MG：数学小道新闻的制造厂．

当《科学美国人》1975年12月的一期被发送到定户时，加利福尼亚的一位热衷的加德纳迷立即翻到“数学游戏”专栏．玛乔莉·赖斯，家住圣迭戈市的一位家庭妇女与五个孩子的母亲，经常是家里阅读他儿子所订杂志的第一位读者．她对七月号上五边形铺砌问题的文章深感兴趣，并且在想“如果克希纳能发现新的五边形铺砌法，那

将多么美妙”．现在，读完了詹姆士新发现的五边形铺法之后，她的兴趣被强烈地诱发出来，于是她决定开始着手，看看她是否还能找到其他新的五边形，用它们来铺砌平面．“我想更好地了解这些诱人模式，看看我是否还能找到其他类型．对我来说它像一个令人愉快的新鲜疑谜，我在考虑怎样才能干得更好．”她的研究一开始就与詹姆士的工作有很大不同，不久之后就对该问题发起大举进攻，并维持了两年之久．

玛乔莉·赖斯在1939年中学毕业，除了高中生必需掌握的一般数学课程之外，从未受过正规数学教育．因此，在她面对挑战，企图发现新的五边形铺砌方法时，她不仅需要搞出自己的进攻方法，而且还得发明她自己的一套记法．她的第一步工作是把马丁·加德纳的有关五边形铺砌的专栏文章提供的全部信息汇总整理一番（见图4）．在干这件事时，她希望能够发现这些五边形及其铺砌法应该满足的共同的关系．“首先，我需要了解这九种类型究竟差别在哪里．我把公式（边角关系）写在一张3×5卡片上，再把10只五边形画在另一张卡片上．接着我用颜色铅笔在五边形内画线以表明公式里的信息，红线表示三个角之和为360°，蓝线表示长度相等的边，黑线表示四个角之和为360°，绿线表示其他信息”．“现在我能看到在类型7与8中，每个顶点都要被一条红线或黑线碰到过两次（图上的挂钩记号算作两次）．如果我在类型1或2中用一根线来表示诸角之和为180°，则每一顶点将被一根线碰到一次．在类型3中继续用此种方法，并用记号↓表示三个相等的内角，这就涉及三个顶点．于是在角B与E之间要作三条直线，每个顶点都要碰上三次．于是我看到，对任一模式来说，五边形的每个顶点都应被线或记号触及相同的次数”．

观察到在铺砌中，五边形的每个顶点必须用上同样次数，这一点是玛乔莉随后研究的关键．她的信息表示法又经进一步精练，只把必要的信息保存下来，其他的都压缩掉（图5）．“我使用的符号五边形很便于作图”．在五边形镶嵌图中将凑集在一点的五边形的各个角顶用线连起来．“这种记法是我开展进一步工作的钥匙．通过角顶的标

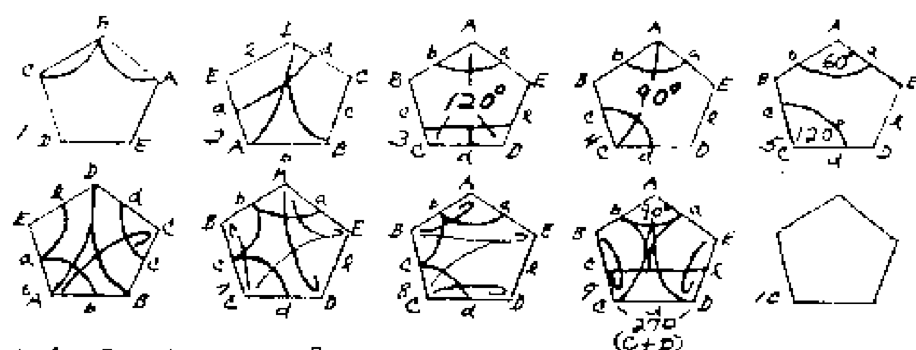


图 4

玛乔莉对 1—8 型可铺满平面的五边形以及詹姆士的发现所作之信息记录.

记(不管从哪里开始),我就有可能用字母组合来把每一张图形记录下来,只需要略略勾上几笔".

玛乔莉通过她自己的方式,发明了一种处理大量信息的办法,而这对于清查各种可能出现的产生五边形铺砌的边角组合必不可少.数学家们为了客观、简洁、明瞭而使用符号,而好的记号必须既能提示,又能定义.玛乔莉的图记法看起来有点类似于象形文字,但它可以记录各个角的可能组合,其简明性是传统数学记法所不可能有的.使用此种记号,就完全不需要去担心由于对五边形的角规定不同的字母而引起的重复现象.

开始时,玛乔莉先把两只五边形拼在一起,然后考虑是否还能把五边形继续放上去以形成平面的一种铺砌法.五边形的各个角顶在铺砌时的会聚,边与边之间的接触,这些信息都用她的符号记录下来.如果这时已能看清楚某种组合是不可能的尝试(不能得到一种

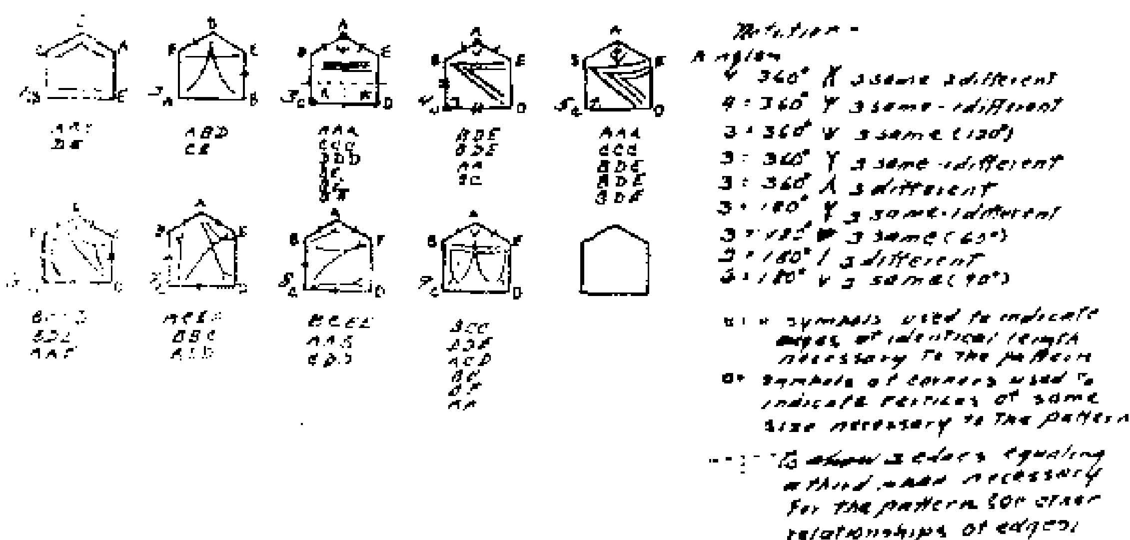


图 5

玛乔莉·赖斯的图形记录法(记下了已知 9 种可铺满平面的五边形信息).这是她全部工作的关键.

铺砌)就立即把它勾销.如果某种铺砌看来是有可能的,她就马上把这种五边形及其铺砌法勾划下来.

为了算得更快,以便于检查新到手的五边形,察看它们的各角相加起来能否得出一种新的五边形铺砌法,玛乔莉把 360°按 18°为单位来划分,在一只小的量角器上标好刻度.用了这种单位,36°角即可记作 2,108°角可记作 6,等等,以此类推."在构筑一个提交试验的五边形时,我通常从其和为 180°的两个角开始,例如 4(72°)与 6(108°),然后再按需要作适当调整.1975 年忙碌的圣诞季节占用了我很多时间,但我只要一有空就去想这个问题,无人时便在厨房的锅台上偷偷画图,一有人来就马上把草图藏起来,因为我不愿意向别人解释我这是在干啥.不久我意识到一些有趣的模式是有铺砌可能性的,但我未作进一步的探索.因为我正在寻找一种新的类型.几星期之后,我终于发现了它".

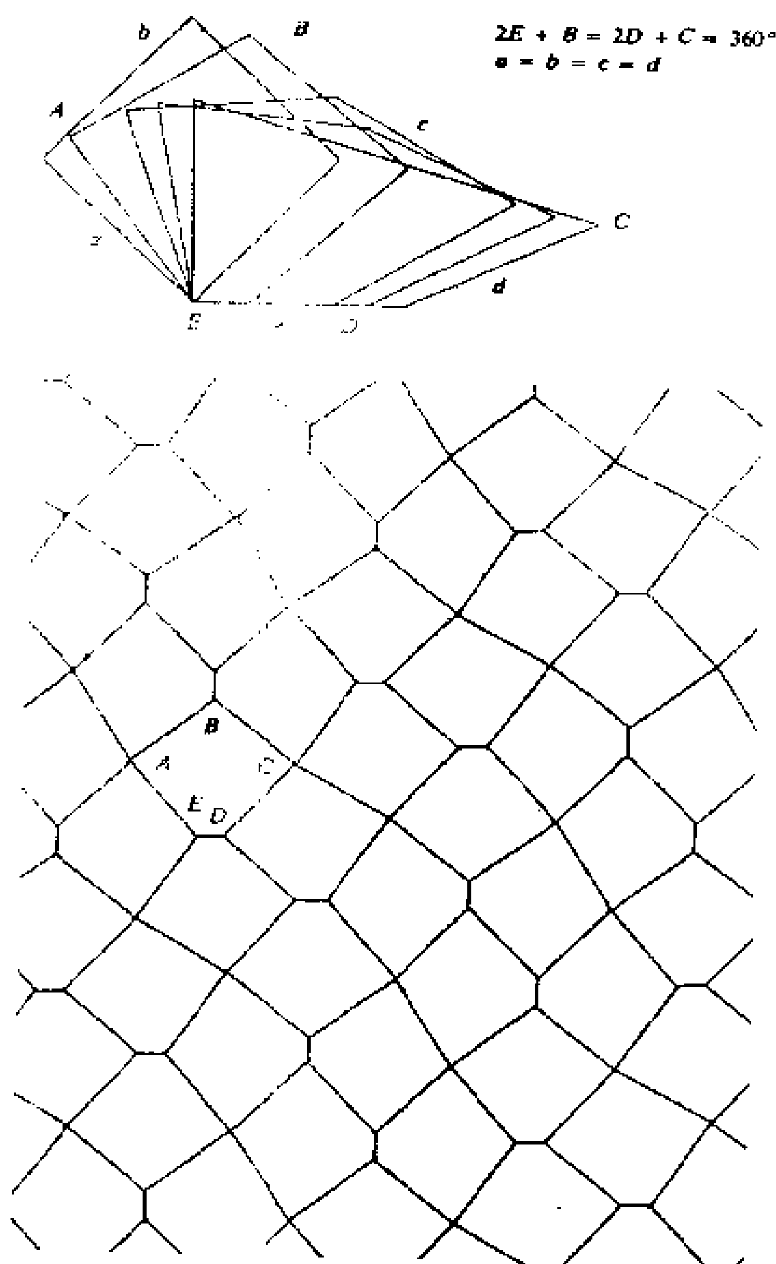


图 6

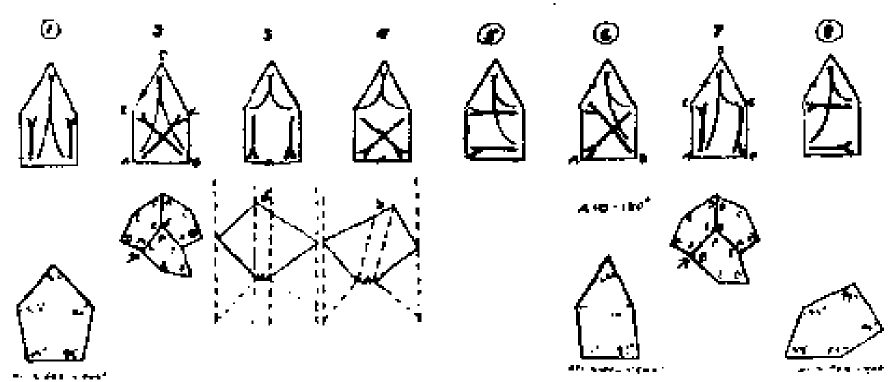
1976 年 2 月玛乔莉·赖斯发现的五边形铺砌图案的一种新类型. 图中给出了五边形所能取的形状之范围以及由此种类型中的一个代表图形所作出之铺砌图案. 由边长成黄金分割比的五边形所拼成的铺砌图案，是埃歇尔型❶美术品"苜蓿叶中的蜜蜂"的基本格子，设计人是赖斯夫人(参看图 16A).

❶ 译者注：埃歇尔是荷兰著名画家，可参看本书中考克塞特教授的文章或译者的另一本著作《数学奇闻录》(1988 年浙江科技出版社刊行).

1976年2月中旬，玛乔莉把她的发现寄给加德纳，信中附有草图，表明了五边形所能取的形状及其伸展范围，以及用这种类型的两个代表所拼成的镶嵌图案(图6)。她写道："这种五边形我确认它不同于您所列举的五边形，虽然它看上去有点类似于第7型和第8型。随信附上的例子中，有一种的边长成黄金分割比，我认为这是一个极为令人喜爱的图案"。加德纳再次立即通过 MG^2 把这个发现迅速告诉一些有兴趣的小团体，其中也包括我和克希纳。它被证明确实是能铺砌平面的五边形清单中应增加的一种。于是克希纳写信给玛乔莉，问她是怎样发现它的，并承认在他的探索中犯了错误，没有把这种五边形作为一种可能类型加以研究。正如大部分读者来信的处置情况那样，在加德纳的专栏里没有报道这个发现，而只是把它归档备查。(作为詹姆士的铺砌法所引起的反响，加德纳也还收到过其他许多来信，受这项设计的启发而产生的美丽图案中，至少有一条被子与一条编织得很好的地毯，请参看彩色插图V.)

当我考察了玛乔莉寄送给加德纳的材料，并把它同克希纳的第7型和第8型进行比较之后，发现似乎这三类五边形(我把她的五边形称为第9型)有可能是更大的一类五边形的特例。于是我作了下列猜想并把它寄给加德纳："任何具有四条等长边和具有满足关系式：$2P+Q=360°, 2R+S=360°$ 的四只不同的角 P、Q、R、S 的五边形，都能铺满平面。"不到两星期，我收到玛乔莉的一封来信，她说已经研究了我的猜想，并证明它是不正确的。"正如下面的记号所表明的，只有8种可能性(只有8种方式可使这种五边形的各角顶点会集在一起并满足关系式 $2P+Q=360°, 2R+S=360°$)。其中第1、5、6、8四种能够铺砌平面，而其他四种看来是不可能的。第6型只能在两邻角之和等于180°时才可以铺砌平面，结果使它变成了第1型。然而它确实给出了两种有趣的装配第1型的特殊类型的方法，请看随信附上之草图"(图7)。

这是我从玛乔莉那里收到的第一封信，我第一次碰到了她的符号以及通过结构来检验可能性的方法。它与数学家们常用的传统方



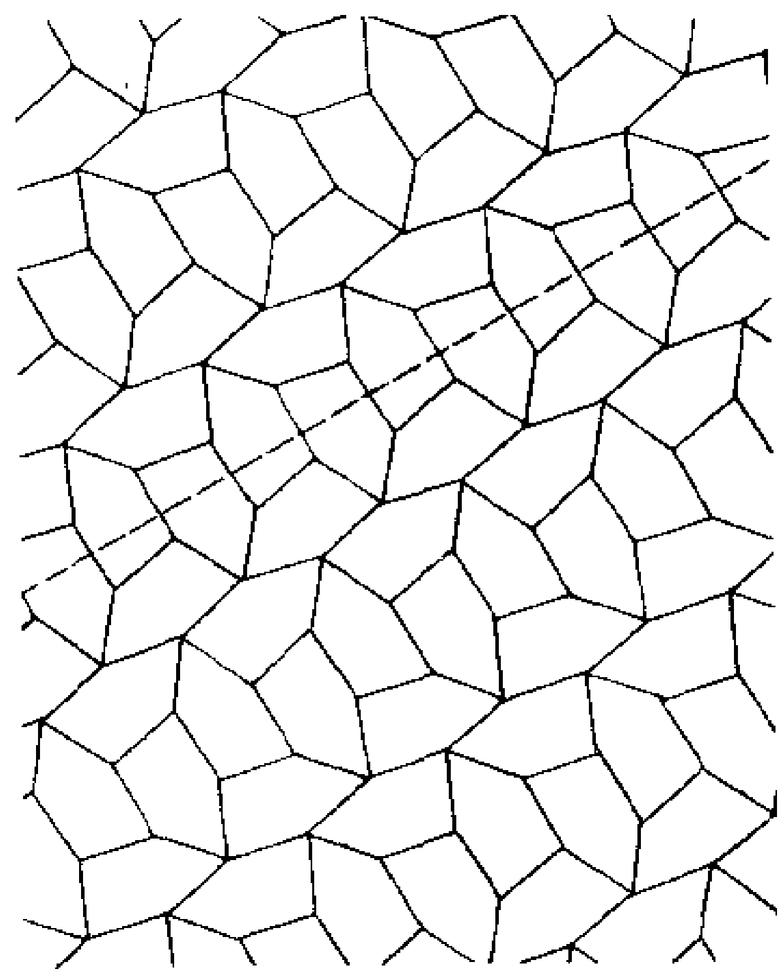


图 7

沙特斯奈德(Schattschneider)猜想不能成立的图形"证明".第6型五边形有4个等长的边,其中有两个邻角之和为180°,此种五边形可由两种不同方式组合为铺砌集成块(请注意穿越虚线时铺砌法的改变).

法差异太大了，以至我眼瞪瞪地注视着图形，企图搞懂她究竟在说些什么，能否证明任何东西. 她所提出的，某些五边形不可能铺满平面的"理由"仅仅是一些小小草图，而不是数学家们所需要的代数或几何论证. 也许是由于它们对她来说是太明显了，所以她在信中没有对其图形记号作出任何解释. 在她的表示法中，类似小鸡脚印的粗线记号⺈表示关系式 $2P+Q=360°$，$2R+S=360°$，这个符号把满足上述关系式的各个角集中聚集在铺砌图案的一个顶点的周围；连接五边形各角顶点的较细的线则表示余下的关系式，指明这些角应当怎样进行拼合. 请回忆一下，五边形的内角和是 540°，连同包括 P、Q、R、S 的两个关系式，就意味着 $Q+S+2T=360°$，这里 T 便是五边形的第五个角. 她认为，五边形的角不能满足更多的关系式，即上述符号已经表示了角之间的所有关系. 这样，她的第 2 和第 7 种情形被排除了，因为(如她的箭头所示)如果五边形要以此种方法去铺砌平面，那就不得不增加另外的关系式. 她的第 3 和第 4 种情形也是办不到的——她的草图说明了这一点. 对我来说，通过代数方法来证明其内角满足上述关系的五边形不可能存在是毫无困难的. 玛乔莉绘出了几个例子(应用第 6 种五边形来铺砌)，在每个例子中两个邻角之和似乎总是 180°，但是这种观察结果并不能算作数学证明. 在试图证明这一观察事实时，我发现假定的角度关系并不能引出这一事实. 后来还是克希纳提供了一个巧妙的证明，表明了他的"推广正弦与余弦定律"的用场. 这桩事情为我们提供了一个突出的例证，它能说明业余爱好者通过其直观想象与观察能力，虽然只利用初等数学工具也能导致正确结论，但要作为无可辩驳的证明，还是需要更高深的数学方法与训练有素的数学教养.

剩下来的四种确实可以铺砌平面，但它早已为人们知晓——她的第 1 与第 8 型分别是克希纳的第 8 和第 7 型. 她的第 5 型是她的新发现(第 9 型)，而第 6 型就是第 1 型. 于是，猜想被彻底处理掉了.

玛乔莉既然已同克希纳和我建立起直接联系，MG^2 就不再起中间作用了，可是，这个问题的进展情况仍然随时向加德纳通报. 玛乔

莉无疑由于来信中对她工作的热情表扬而深受鼓舞，但是继续吸引住她的还是问题本身．虽然她忙于家务，但她不断地探索这个难题，挤出断断续续的时间来复制铺砌草图，反复推敲各种可能性．这个题目犹如放在家中闲置空房里的、业已拼出一部分的拼板游戏——专心思索一阵子，得到了一个小小的满足，再把它搁一搁，但并没有忘记它，它会再度引诱你去推敲，加上几块组件，看一看拼好的图案．

当我接受《数学杂志》约稿，请我写一篇五边形铺砌问题专稿之后，我就要求玛乔莉把她在这个问题上所做过的一切工作写信告诉我，并通报任何新结果．1976年3月，我收到了她用以进行研究的、经过整理的分析方法．图形表明她把五边形分成几组，每一组对应着一个五边形所应满足的内角关系．这些内角关系式（并且仅是这些关系式）要在由这个五边形拼成的铺砌图案中得到反映．角怎样拼合也就促使人们去考虑在五边形的边长方面应满足什么关系．第一组称为*P*1（第一类五边形的缩写），研究的是每个角在铺砌图案的一个顶点周围只“使用”一次的五边形．于是，三个不同的内角拼合起来，它们的和就是360°，两个剩下来的内角拼合起来，它们的和便是180°．只有两种五边形（第1型与第2型）属于这一类（图8）．余下来的12个组（在标题“*P*2”的下面编为1至12号），研究的是每个角在铺砌图案的不同类型的顶点处总共“使用”过两次的五边形．在这张清单中的第12组就是玛乔莉为答复我的猜想而在早些时候写信告诉过我的东西．在这张整理过的清单（图9）中，玛乔莉解释了“要进行三步

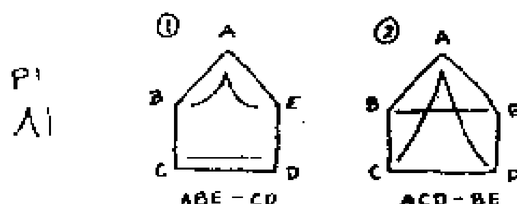


图　8

类五边形的符号表示法．其中五边形的每个内角在铺砌图案的每个顶点周围只“使用”过一次．

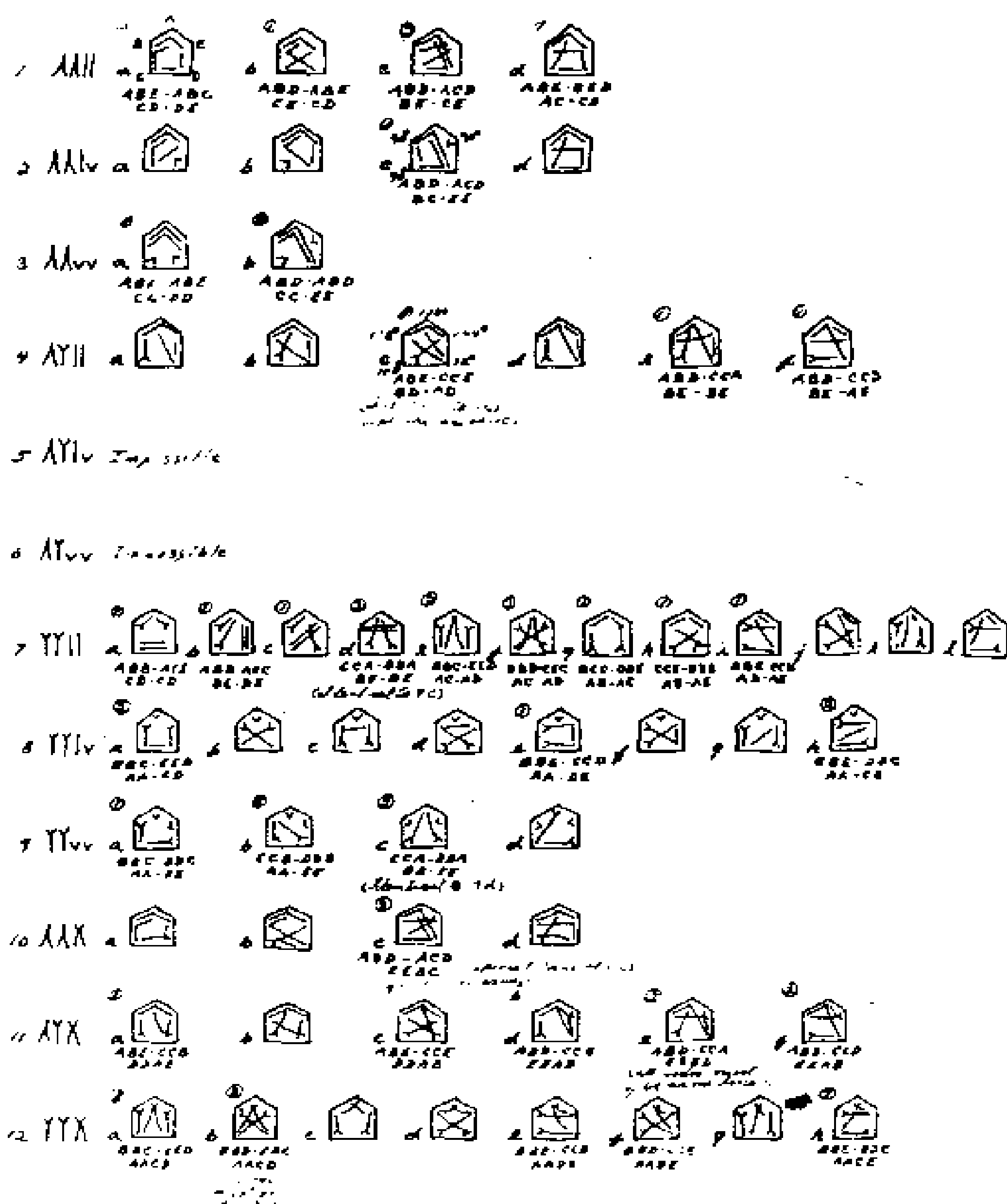


图 9

形成 $P2$ 模式的 12 组五边形. 这类五边形的每个内角在铺砌图中被"使用"上二次. 对每一个获得成功的内角组合方式,在图 9A 与图 9B 中都分别给出了具体的铺砌草图. 这张单子以及各种铺砌法汇集是玛乔莉·赖斯所作出的第二个贡献.

试验，第一步要判明这一组本身是否可能（第5和第6组就不可能）．然后对某个组中的每个五边形画出草图，以检查适当的内角能否拼合．如果能够拼合的话，那么五边形的哪几条边应该相等就变得很清楚．最后一步是，这些角能否变成特殊的角——如果能，那就可以成功地铺砌"．如果照它的安排方式拼合不起来，玛乔莉便在图旁记上一个"不"字；如果可以铺砌，她就记上一个从克希纳的单子中取来的型号．不仅如此，对每种可以铺砌平面的五边形，她还具体给出了实际铺砌图案（图9A，9B）．在她的铺砌清单中，第1、2、4、6、7、8、9型都曾出现过．她一共有了26种不同铺砌法．其中有一些是新出现的．第3和第5型以及詹姆士的铺砌法（我把它称为第10型）在这张单子中未曾出现是不足为怪的，因为这些类型的铺砌图中，各个顶点处的内角关系式都能满足时，则五边形的每个内角势将被用上三次．

在收到这一信息数星期后，玛乔莉送来一张更大的清单．"在重新考察了以前我寄给您的$P2$模式之后，我发现遗漏了一些，于是我再次仔细地作了检查，谨将修正清单连同其实例一并寄上．"她的修正清单中罗列的虽然也是同样的12组，但却考虑了更多的情况．迄今她已找到35种五边形以及导致平面铺砌的内角关系．某些内角组合能够产生两种以上不同铺砌法，因而这张清单一共收入了45幅铺砌草图．虽然没有发现新的铺砌类型，但可以铺砌的五边形总数大大地扩展了．来信表明她已在研究$P3$模式，即内角关系将把五边形的每个内角使用上三次．"其中的大多数可以迅速地判明为不可能，所以把余下来的统统检查一遍所需的工作量并不过于庞大……在它们中间，可铺砌的五边形有第3型、第5型以及詹姆士先生新发现的一种类型（第10型）．"

1976年10月，我从玛乔莉那里又收到另一封胀鼓鼓的信件．她又送来了一份新的清单，其中收入了迄今她发现的全部五边形镶嵌法——总共58种．她已重新编排了清单．这一次她将五边形（及其相应的铺砌图案）分为12组．凡具有同样的几条等边关系的五边形都归入一组．长达六页的铺砌图案充分说明了她的研究工作的彻底性．

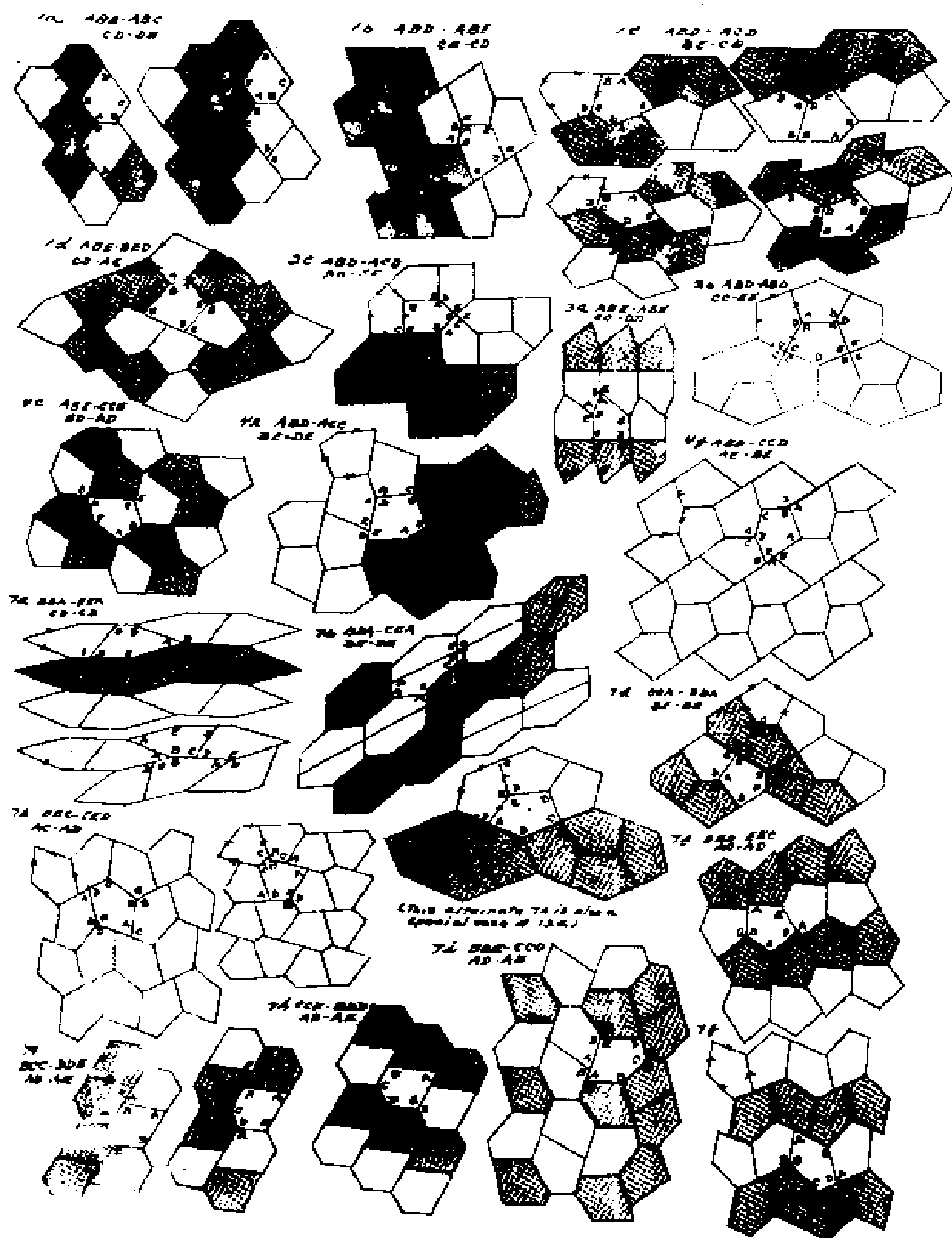


图 9A

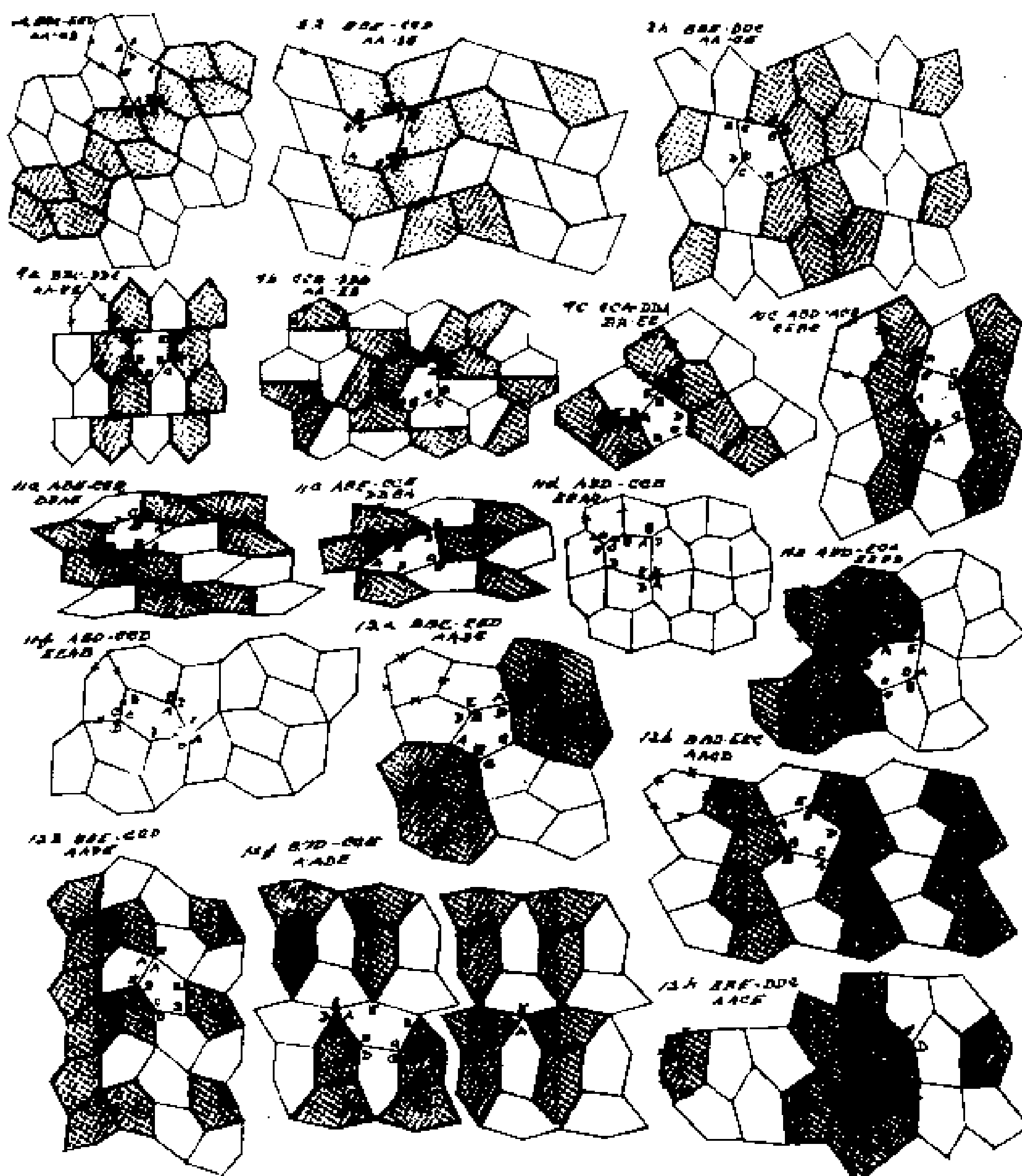


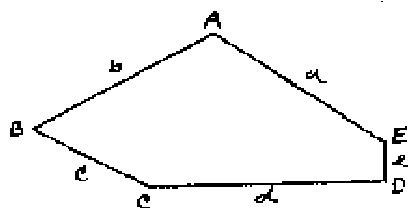
图 9B

十种类型五边形中的每一个都给出了具体铺砌法，其中出现了新的拼合法. 她用一句话作为这封长信的结尾："到此为止，才疏学浅的我已经走到了尽头. 我已经是露了底. 也许仍然有一些东西没有被我发现."

十一月中旬，我寄送玛乔莉一份预印件，它是勃兰古·格隆鲍姆

(Branko Grünbaum)和杰弗里·谢泼德(Geoffrey Shephard)所写的论文,显示了利用第 1 至第 5 型五边形的 24 种块传递式的平面铺砌法. 另外,信中还附有我为《数学杂志》撰写的五边形铺砌问题的文章初稿. 它是我在俄亥俄州牛津市迈阿密大学召开的游戏数学学术会议上所作之讲演的扩充改写稿. 约翰·H·康威(John H. Conway)❶出席了这个会议,他对该问题以及詹姆士与赖斯的贡献表现出莫大兴趣. 他承认自己也曾想把所有的可以铺砌平面的五边形统统找出来,但由于耗时太多而终于放弃. 在我这篇文章的末尾,我提出了几个很自然的问题:"采用边与边拼合方式的、可以铺满平面的五边形清单是否已经完备了?""我们能否找到所有的等边五边形以铺砌平面?"

玛乔莉不可能不理睬这些问题. 1976 年 12 月,我又收到另一封信. "我不打算在五边形问题上再花时间了,可是它们并不是很容易搁到一边去的. "这一次,她已经同我在论文中所提到的问题交过锋. 在回答末一个问题时,她已画出了所有用等边的五边形作出的铺砌图案以及铺砌法所要求满足的内角关系式. 她也触摸过第一个问题.



$D=90°$

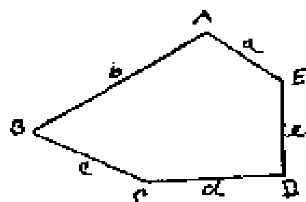
$B+E=180°$

$A+A+E=360°$

$C+C+B=360°$

$a=b$

$e+e+c=b$



$D=90°$

$B+E=180°$

$A+A+E=360°$

$C+C+B=360°$

$a+c=b$

$e+e=b$

(angles of the 2 pentagons · 117°-54°-153°-90°-126°)

❶ 译者注：英国剑桥大学的著名代数学家,在趣味数学方面也有很多建树.

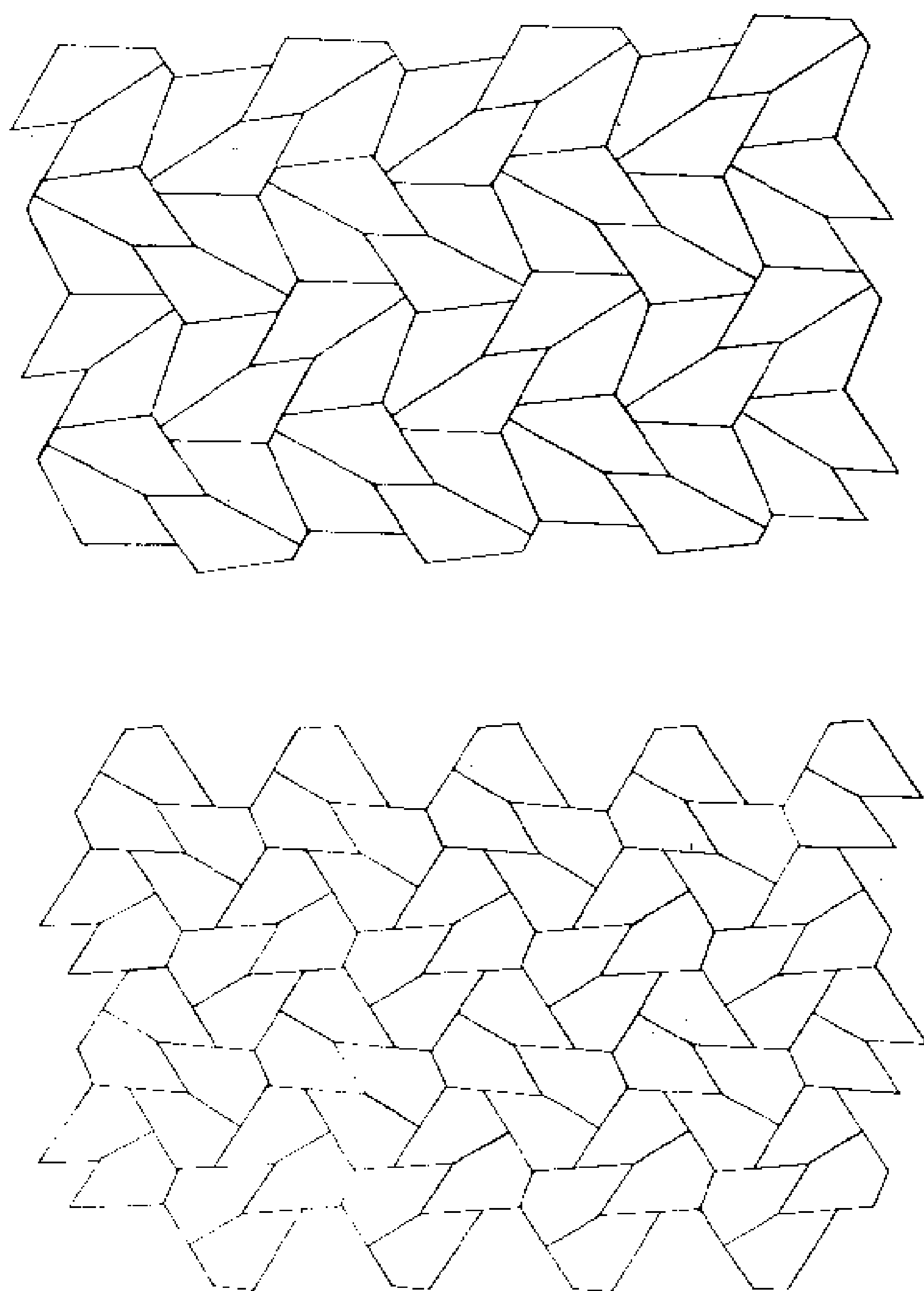


图 10

玛乔莉・赖斯在 1976 年 12 月所发现的第 11 型与第 12 型五边形，用它们可以铺满平面.

在那篇文章中，我解释过五边形的“集成块推移”铺砌法. 因为我注意到，在所有的新发现中(其中有克希纳的、詹姆士的以及她自己的)，并不是用一个个孤立的五边形来铺砌，而是两个或三个五边形结成集成块来铺砌. 这个概念，对玛乔莉来说是新鲜的，于是她重新检查了她所有的铺砌法，其中第一个图案便是用两个五边形组成一个集成块，她又注意到“它们中间的绝大多数是把四个五边形集成为 2 个六边形，而用后者铺满平面，则总共有 6 种办法”. 集中精力研究一个集成块怎样划分为四个全等的五边形，这使她又发现了一些新奇的边对边的拼合法. 两星期后(1976 年 12 月 27 日 —— 圣诞季节似乎是她最有创造性的时刻!)终于传来令人兴奋的消息:“我一直在利用这概念进行探索并有了一些新模式. 令我振奋的是，2 种新类型是具有紧密联系的.”她真的发现了第 11 型和第 12 型. 相应的铺砌图案十分引人注目(图 10).

新类型的发现在她对 2-块集成推移拼合所作的方法论分析中是个意外收获. 她已发现双六角形可按 9 种不同方法划分为 4 个全等的五边形(图 11)，这些划分及多种铺砌大大丰富了镶嵌图案，使五边形的不同铺砌法(这些铺砌法都是“2-块集成推移式”的)总数超过五十种. 整个春天她坚持追随这个新概念进行工作，发现了 3-块集成甚至 4-块集成的五边形铺砌平面图案.

原先由 MG[2] 传播信息的五边形铺砌问题现已日益茁壮成长，有关信息被传播到三大洲. 另一群业余爱好者——澳大利亚新南威尔士州的几位十一年级中学生花了一星期功夫研究，发现可以铺砌的凸等边五边形及其具体铺砌方式. 在他们的两位老师乔治・捷克斯(George Szekeres)与迈克尔・希尔锡洪(Michael Hirschhorn)的指导下取得了很好进展. 一种特殊的等边五边形能用多种不同方式铺砌平面(它就是玛乔莉・赖斯第一次和我通信时提到的“第六种”五边形). 希尔锡洪发现，利用这种五边形可以得出许多很不寻常的铺砌图案，其中也包括两种美丽的中心辐射式铺砌图案，它们正好具有六重旋转对称性(图 12).

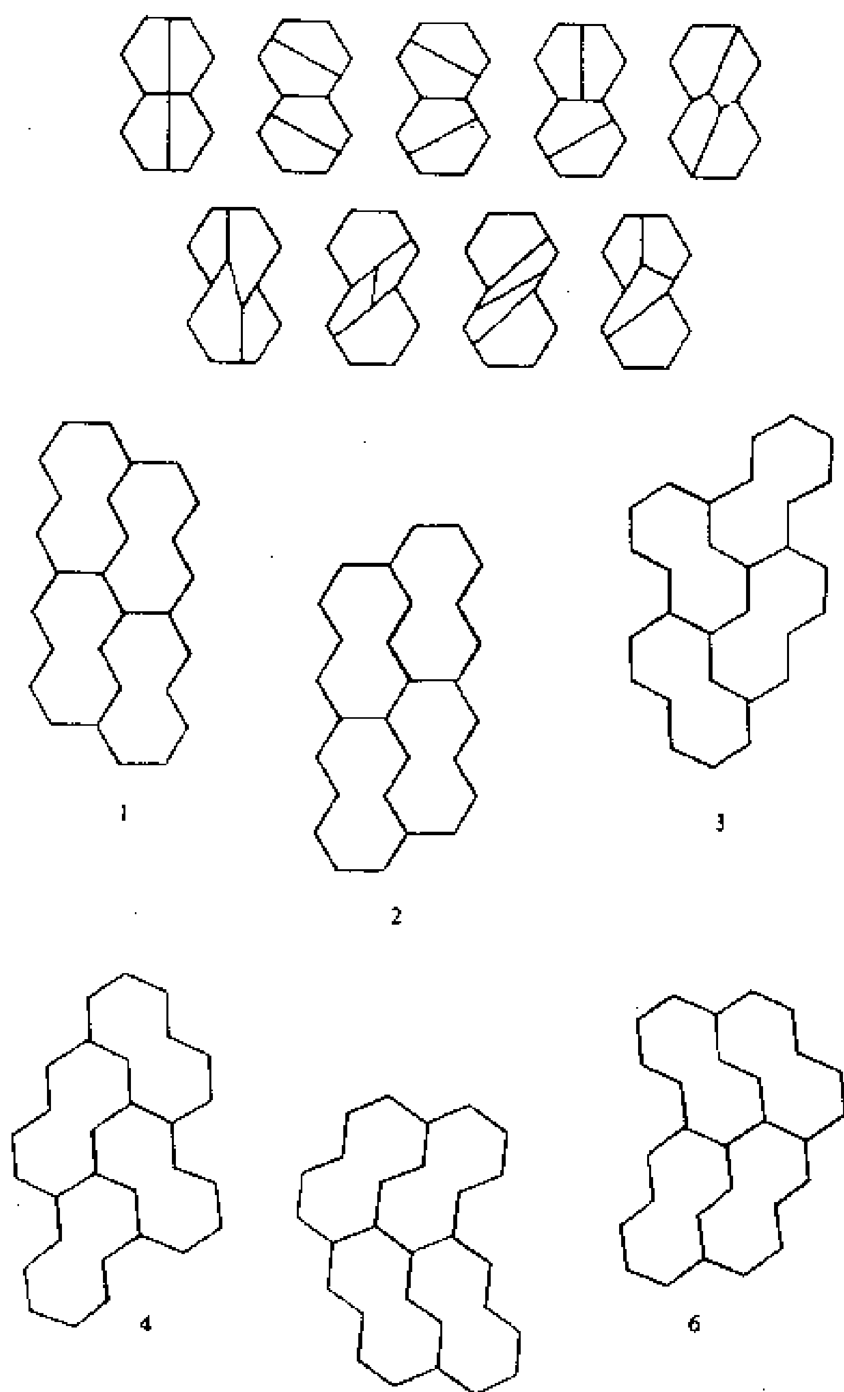


图 11

研究"双六边形"集成块(这里用符号表示)怎样划分为四个全等五边形. 探讨这种双六边形块铺砌平面的方式,其结果产生了一批五边形的新铺砌图案.

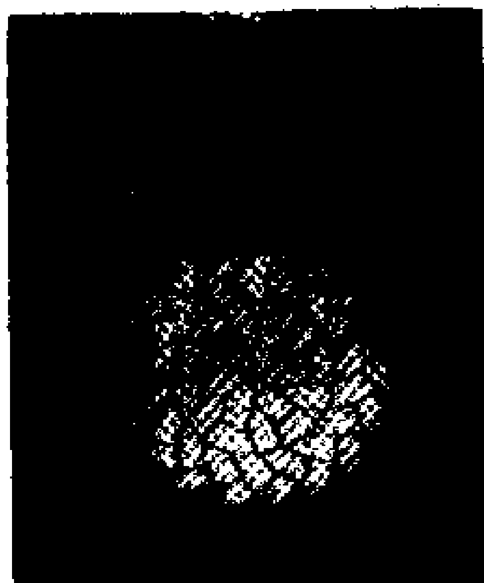


图 12

迈克尔·希尔锡洪的中心辐射式铺砌图案，所用元件是一个等边的五边形．它已被加工精制成银质挂件．

1977年夏季，我向玛乔莉提供了勃兰古·格隆包姆与杰弗里·谢泼德的论文"平面的81种等面铺砌"．论文中附有绘制精巧的81种不同铺砌模式．我觉得玛乔莉有可能从它们中间找出铺砌集成块可以进一步分解为若干个全等五边形的办法，从而得出一些新的五边形铺砌图案．整个秋季，她继续进行着以往的工作，并以一种远非我所能企及的熟练技巧运用着格隆包姆与谢泼德的论文．"格隆包姆与谢泼德的等面铺砌图真是太有趣了．我已把它摘录成4页篇幅以便使用"．事实上，她确实已把格隆包姆一谢泼德改制为符号标志的铺砌图，其中只表明铺砌的拓扑网络以及铺砌对称群的作用(图13)．利用这些资料，她再次分析研究了她以前的各种铺砌图并且又发现了一些新的铺砌图．

当她把一只装有大量工作成果的特大信封寄给我时，那是1977年的圣诞季节．那可是再度出现的圣诞奇迹．"正好两星期之前，这种

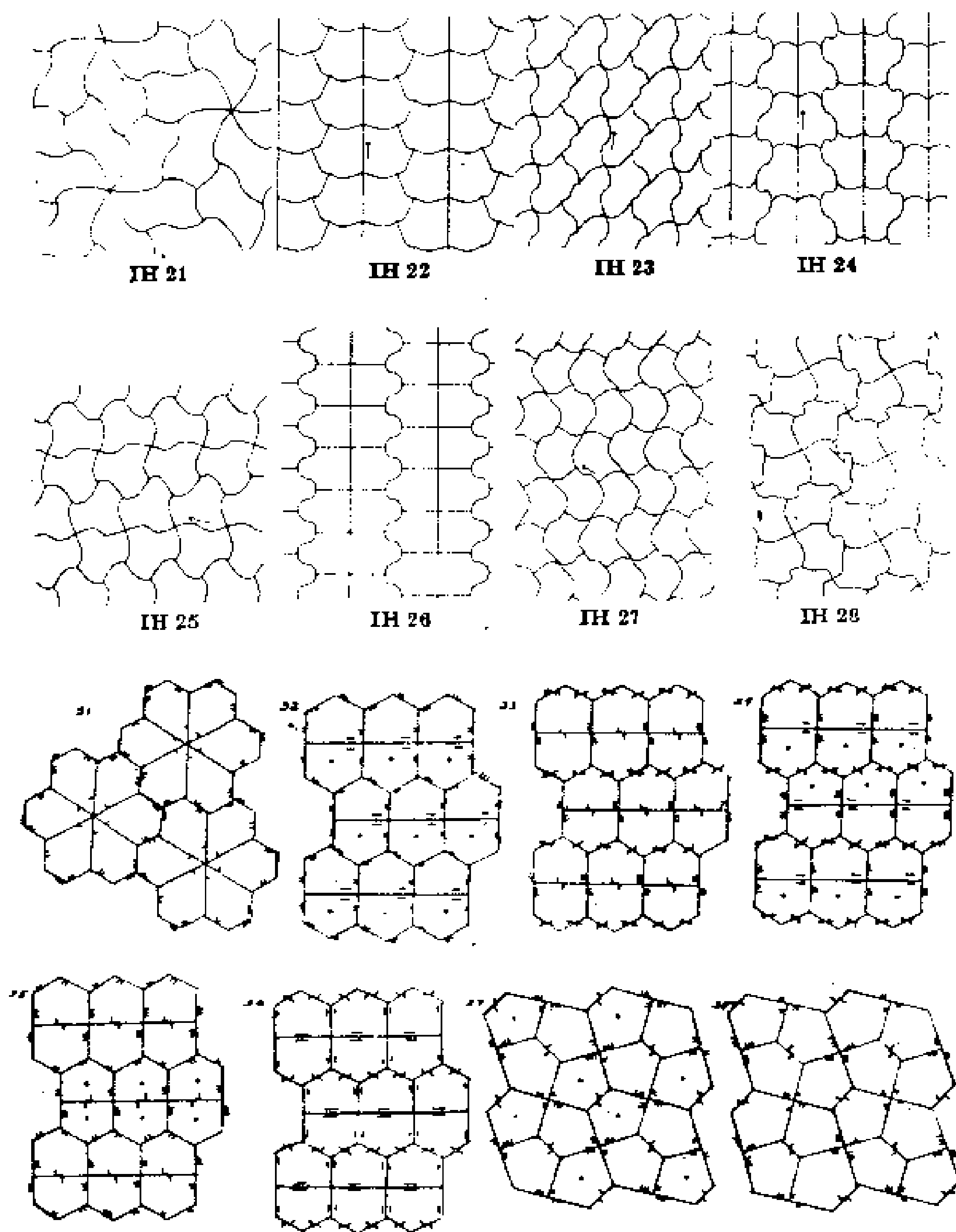
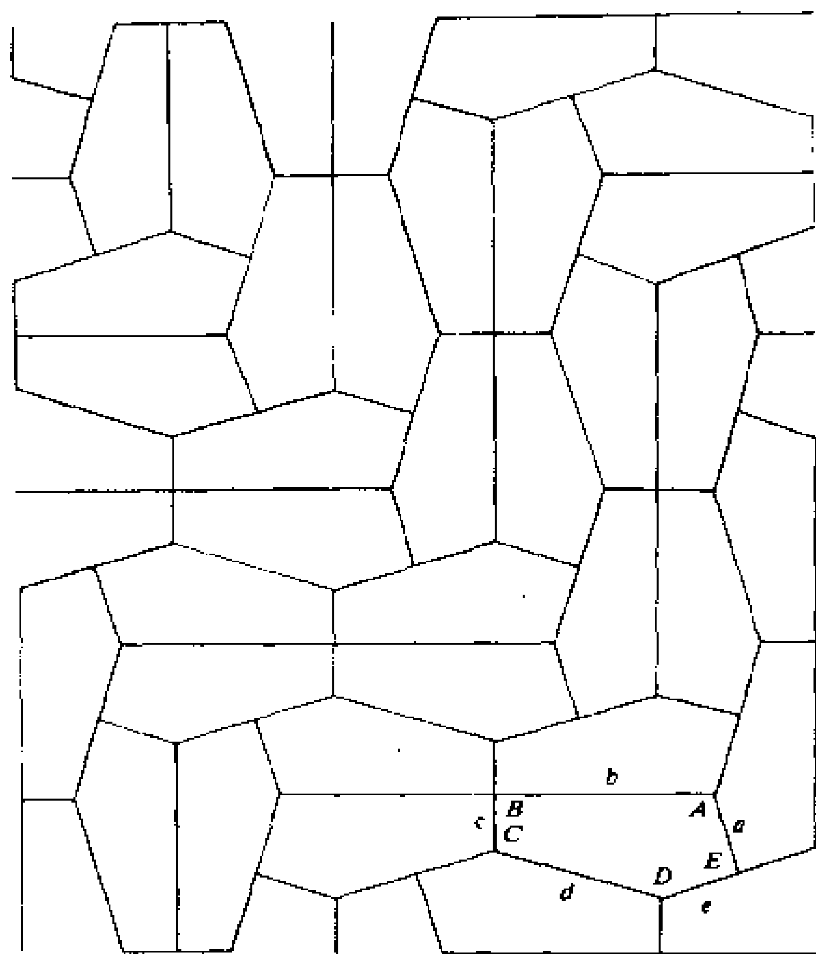


图 13

（上半部）出现在论文"平面上的81种等面铺砌"里的几种形态铺砌图案．作者为勃兰古·格隆包姆与杰弗里·谢波德（见剑桥哲学会数学论文汇编，1977年9月号，第190－191页）．（下半部）上述铺砌图案经玛乔莉·赖斯改绘后的符号形式．

新型的五角形突然涌现(我认为不会再有了).这个东西(见图14)类似于第4型,有着两只对角各为90°,然而对边长的要求却与以前的不一样."作为这种新的、第13型五边形的铺砌法图解显示了一种非常有趣的模式——连锁蝴蝶结,由四只五边形结合而成.此时,我正好接到我那篇文章的校样,因此就有可能在正式刊出之前把这一最新发现增补进去.(发表在《数学杂志》1978年1月号这一期上的论文,对能铺满平面的五边形作了极充分的讨论,备述各项数学细节.)



$B = E = 90°$
$2A + D = 360°$
$2C + D = 360°$
$a = e$
$a + e = d$

图 14

玛乔莉·赖斯在1977年12月所发现的第13型可供铺砌的五边形.

玛乔莉在这个问题上的研究工作依然没有告一段落，她决心作出尝试，要看一看她能否把所有可以铺砌平面的五边形统统找出来. 虽然她的证明尝试是不完备的，可是她对所有的 2-块与 3-块模式所作的、透彻的组合检查还是使剩下的工作量大大减轻了. 也许，当这一故事被印出之时，是否仍有其他可铺砌平面的五边形存在的问题将可作出回答. 希尔锡洪信心十足地认为，能铺砌平面的一切等边凸五边形都已找到了(在《数学杂志》的论文里有所说明). 他的办法是用计算机证明，采取的手段是排除各种可能出现的角关系式.

何以玛乔莉竟能如此孜孜不倦地盯牢一个问题？她不是专门培训来对付它的，更不是为了挣钱，而是在耐心与坚持研究上明显地获得了个人的满足. 毫无疑问，她的经历表明她也像别的业余爱好者那样，深受加德纳著作的启发.

她于 1923 年出生在佛罗里达州的圣彼得堡，是家庭里的老大. 五岁时，她就读于花园谷学校，那是一所只有一间课堂的乡村小学，从一年级到八年级，总共大约有两打学生. “我母亲希望我有一个良好的开端. 她在家庭里就把我教得很好，因此我被插到二年级读书”. 她是一个腼腆的孩子，喜爱读书. “很容易被书中的情节或我自己的白日梦吸引住，几乎忘掉了周围的世界”. 她在学校里成绩不错. “算术对我来说很容易，我希望发现躲在方法背后的理由”. “我对自然界的色彩、模式与设计深感兴趣，梦想有朝一日成为一位艺术家……”在这所学校的后几年，她在两位很好的老师基赛(Keasey)小姐与提蒙斯(Timmons)小姐那里得益匪浅，她们弥补了一所简陋的乡村小学的种种缺陷. ”

“在我读 6 年级或 7 年级时，我们的老师有一天指着油画镜框，向我们介绍了黄金分割比. 它几乎马上抓住了我的想象力. 这虽是一桩偶发事件，我却永远不会忘记. 后来几年，我阅读了各种领域的知识，特别感兴趣的是建筑学以及建筑设计师们的思想，譬如像布克明斯特 · 富勒(Buckminster Fuller)等大师. 后来，我在阅读时再次遇到黄金分割，思索了它在绘画与设计中的应用. 对我特别有帮助、有启

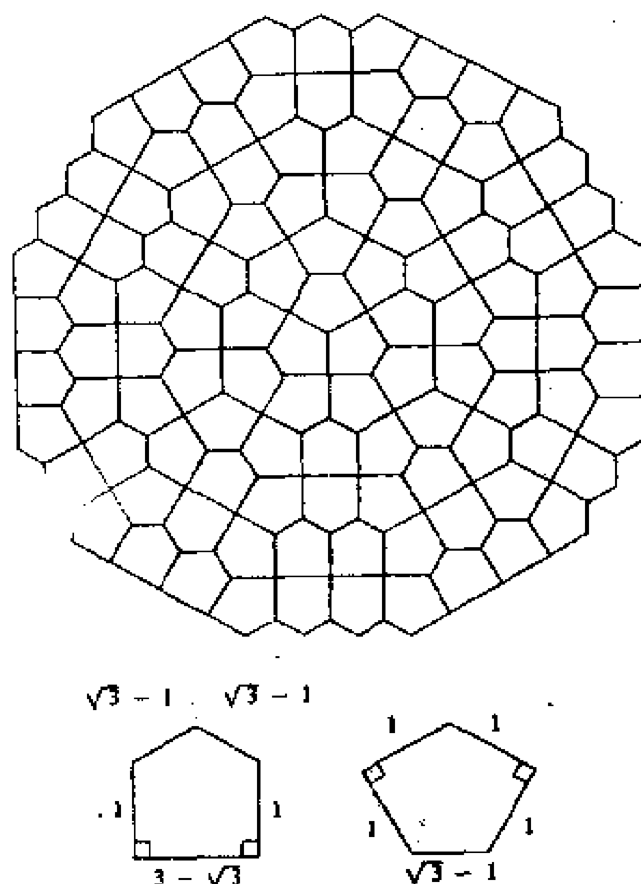


图 15

玛乔莉·赖斯设计的一个美观的对称铺砌图案，自中心向外辐射，扩及到全平面. 其组成元件为两种五边形，各有两只直角与三只 120°角.

发的书是马提拉·廓卡(Matila Ghyka)所写的《艺术与生活中的几何学》."玛乔莉对艺术的兴趣经久不衰，她变得对织品设计以及 M.C. 埃歇尔的作品特别感兴趣. 在她探索五边形问题及其铺砌法时，她制作了一些美观的几何图案以及富于想象力的、埃歇尔式的作品(图 15 及图 16).

在读高中时，她们一家搬迁到了佛罗里达州的松堡，靠近渥兰多市. 玛乔莉在渥兰多高级中学里学了速记与打字，准备就业，但这两种技能都未学好. 她很后悔，除了必需的一般课程之外，没有进一步

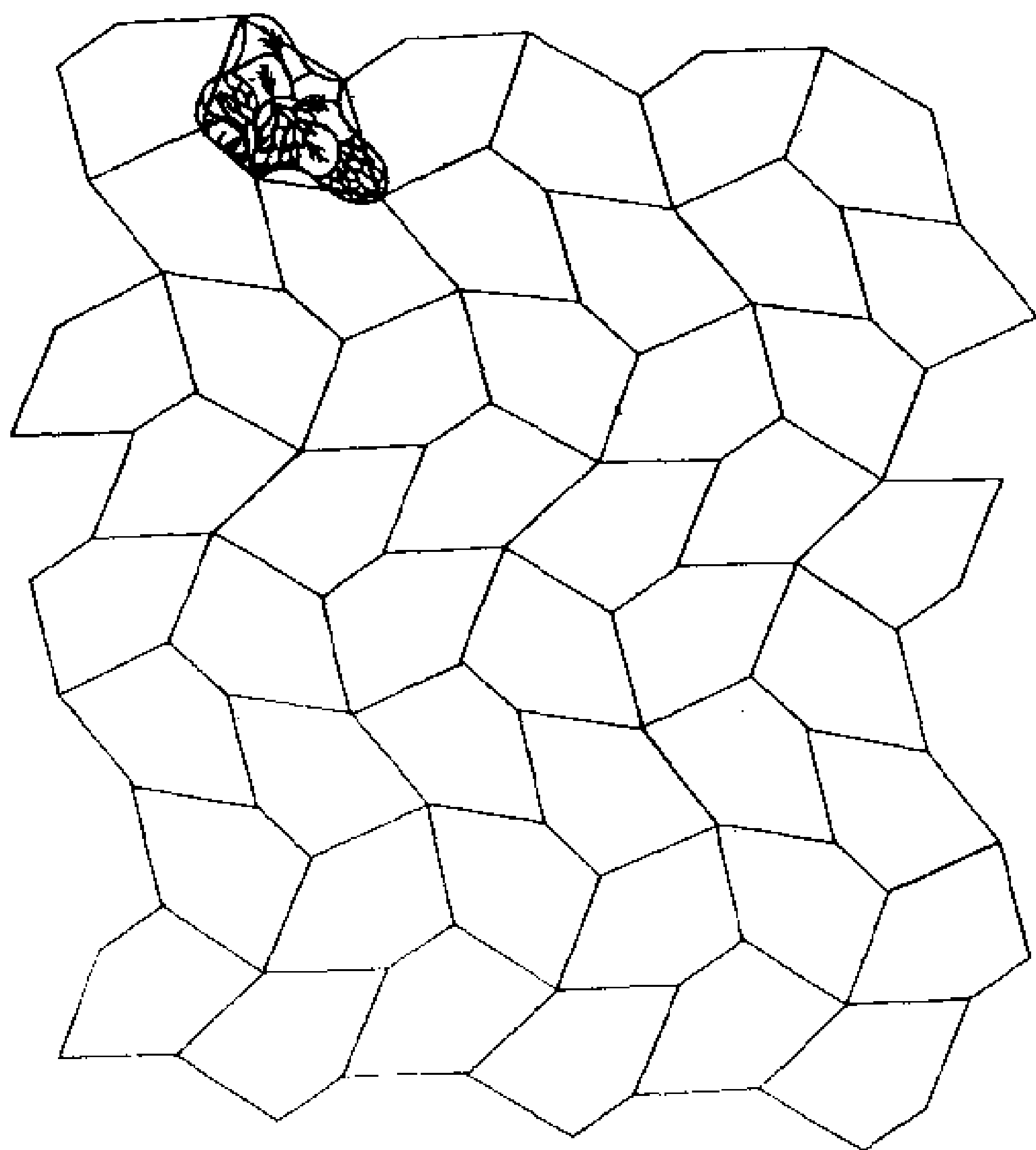


图　16A

"苜蓿叶中的蜜蜂"的构图基本格子(见彩色插图 I).

图　16

玛乔莉·赖斯所设计的三幅仿埃歇尔作品,其几何框架来自她发现的五边形新型铺砌图.(M.C.埃歇尔曾利用过一个众所周知的五边形铺砌图作为他的某些平面镶嵌图的奠基格子.)

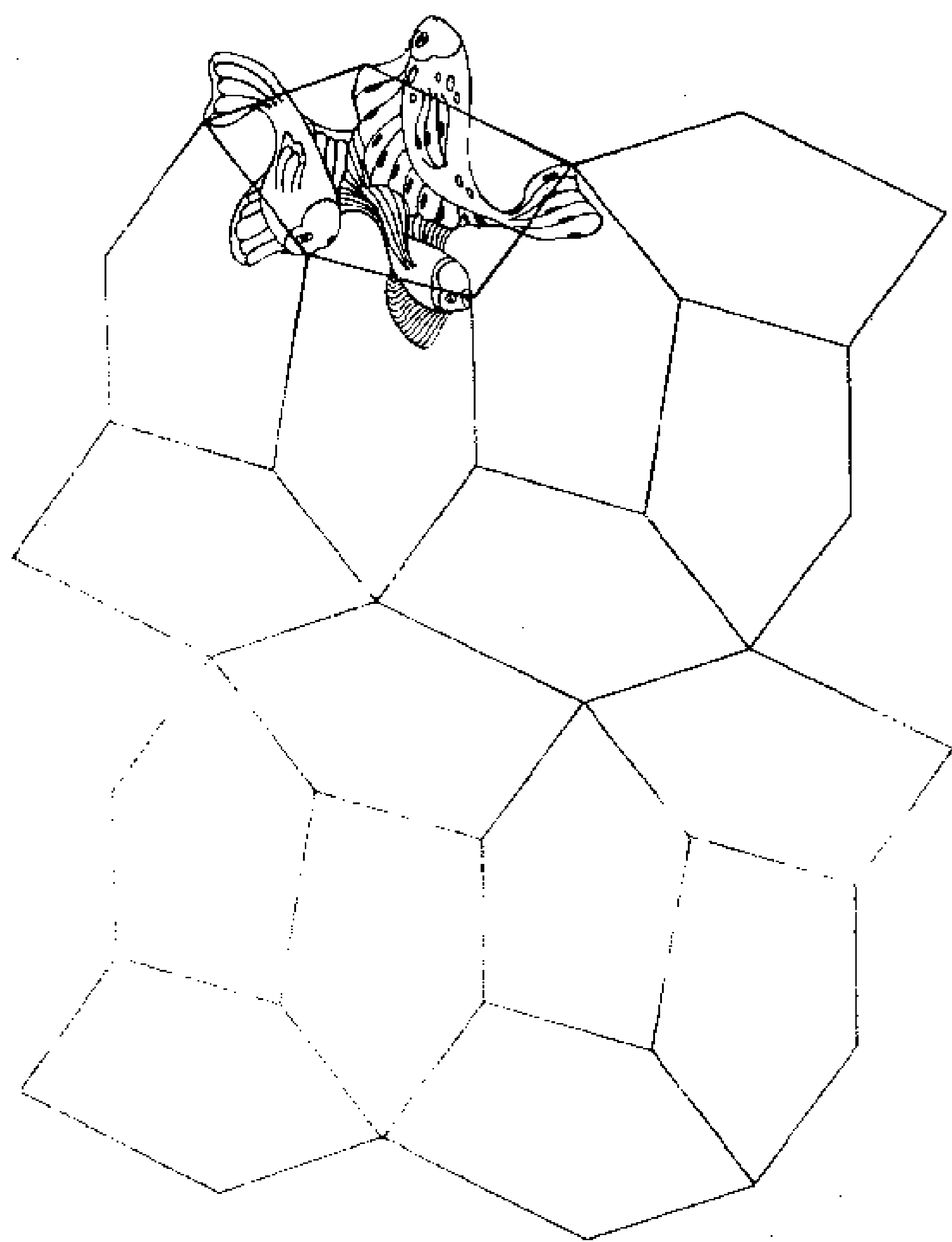


图 16B

"鱼"的构图基本格子(参看彩色插图Ⅱ).

学习数学. 16 岁高中毕业后她先是受雇于一家洗衣房，后来又转到一家小印刷所，直至 1945 年与吉尔伯特·赖斯(Gilbert Rice)结婚. “在那些年头，我经常跑公共图书馆，读了许多有关科学、心理学与学校里没有学过的其他课程，并且开始了应用艺术的函授学习…….”在赖斯夫妇的长子出生后，她们全家迁至加利福尼亚州圣迭戈市. “对我们来说，这些年头是忙得不可开交的”——这句话确非谦词. 赖斯夫妻一共生了五个孩子，吉尔伯特也开起了他的铸字店. 玛乔莉是

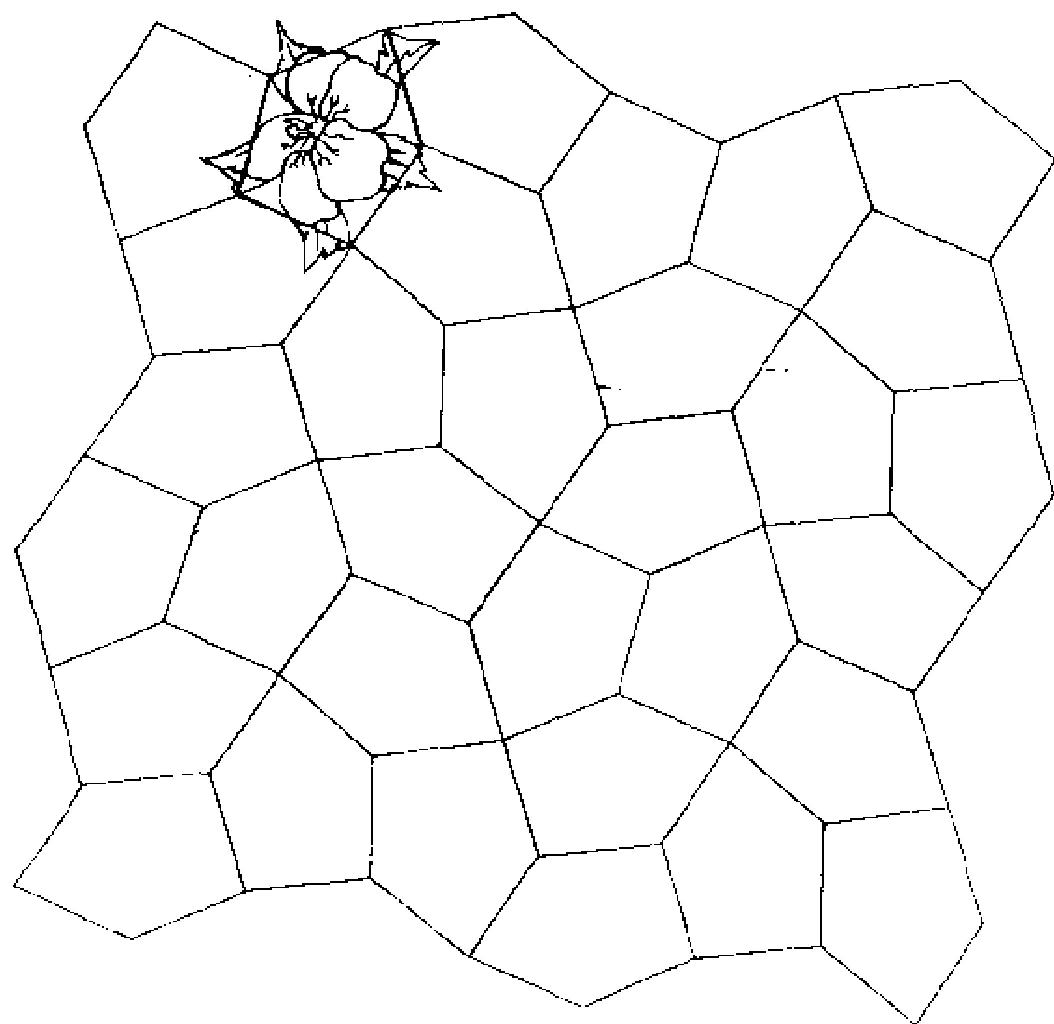


图 16C

“木槿花”的构图基本格子(参看彩色插图Ⅲ).

被她的儿子们拉进数学里的."当我的大儿子戴维(David)进初中时,'新数学运动'兴起,开始进入学堂……我想读他的课本并与之保持同步,他也鼓励我这样做,但这些日子家务繁忙,不久我就感到自己落后了一大截,放弃了这个念头.尽管如此,我对数学课的兴趣时断时续,我经常能通过一些不正规的办法找出他的问题的答案.因为我不知道正当的步骤究竟是什么.他同我一起做他在班上学到的数学游戏,例如六角形游戏 Hex,三维吃井字❶……"

"我喜欢各种各样的谜题,十字谜,拼板游戏,数学谜题与游戏,多年来买进一批数学游戏书.尤其爱好其中几何味道特别浓的东西.因此,当我儿子订阅的《科学美国人》杂志到达时,我马上就翻到马丁·加德纳的'数学游戏'专栏".她对这类游戏的专注程度,对形状、比例与图案设计的敏锐观察力在下面的一段话里作了充分表达,那是她最近一次旅游生活的记述:"1974 年 11 月,我和我的丈夫开始作环球旅行……我对风格特异、为我们不熟悉的各种设计、比例表现了莫大的兴趣与好奇.不论到哪里,我总是要把房屋、门窗、空间分隔、窗格图案……的各种比例记入笔记本.我特别欣赏那些令人愉快、色彩绮丽的纺织品设计图案,此种大胆的设计在加纳和尼日利亚等非洲国家经常可见"."在这次旅行生活中,我们尽量减轻行装……但我还是偷偷藏进一本小册子,L. H. Longley-Cook 出版的《把这个题目解出来! 105 个诱人的动脑筋问题》,当我们长时间等候车、船时,想想这些问题,时间就会过得飞快……."

这些业余爱好者的长处是其心理与精神状态——强烈的刨根问底精神以及对他们见到的一切事物的敏锐观察力.任何正规教育造就不了这些素质.把这些"业余爱好者"与"专业数学工作者"分隔开来的仅仅是这些人没有拿到数学学位.但是,他们的不屈不挠的好奇心与机敏巧妙的办法使他们成为真正的数学家.正是马丁·加德纳唤醒了许多这样的数学家.

---

❶ 译者注:一种中、外儿童都会做的儿童游戏在三维空间的推广.

平面铺砌的若干问题

● 华盛顿大学

□ 勃兰古·格隆包姆(Branko Grünbaum)

● 东安格里亚大学

□ 杰弗里·谢泼德(G. C. Shephard)

虽然铺砌艺术几乎同人类历史一样悠久,然而铺砌的科学却似乎一直不受重视,直至最近才有所改变.

中世纪的灿烂艺术在一些伊斯兰教寺院与别的萨拉森式建筑的引人注目的铺砌图案中提供了实例,其中14世纪的西班牙王宫——众所周知的阿尔亨布拉宫(Alhambra)尤为著名(见图1与图2).对阿尔亨布拉宫的参访无疑引起了荷兰艺术家M·C·埃歇尔的创作热情,从而产生了一些著名的杰作(见图3).

从数学角度对铺砌问题进行探讨首见于德国天文学家约翰尼斯·开普勒(Johannes Kepler,1571—1630)(见图4)所著、1619年出版的《宇宙和声》一书中(见图5).开普勒的几何发现被他自己在天文与物理方面的伟大成就所掩盖,难以置信地被人们遗忘了将近三百年之久.除了开普勒的这部著作之外,直到19世纪末年,在平面铺砌问题上的著作少得可怜,因而,铺砌的科学研究(我们指的是其数学性质的活跃研究)仅仅只有一百年的历史.在本世纪,已有大量资料陆续发表出来,其中极大多数来自结晶学家、工程技术人员以及其他非专业的数学工作者.这一课题中仍然充斥着许多悬而未决的问题.

我们将在下文对其中的一些问题加以探讨.



图 1

西班牙格拉纳达市阿尔罕布拉宫的内景一瞥. 与其他萨拉森建筑物一样,这一14世纪的王宫以其大量精美绝伦的装饰性铺砌图案而蜚声于世. 附图2将给出阿尔罕布拉宫的一部分铺砌图案.

最近一个时期,由于人们认识到几何在教学上的重要性,铺砌问题已被认为是大、中学校的一桩合适的数学课外活动. 人们不无兴趣地注意到,以数学教师为主要读者对象的一些数学杂志似乎已为这一课题留出了越来越多的篇幅. 此种兴趣的复苏,一部分原因自应归功于马丁·加德纳在《科学美国人》杂志上所撰写的动人文章. 他经常报道新发现,不断刺激了这方面读者的兴趣. 下文我们将有可能提到一些这方面的论文,并殷切希望数学界对马丁·加德纳这位最出色的传播者应具有明确的感戴心情.

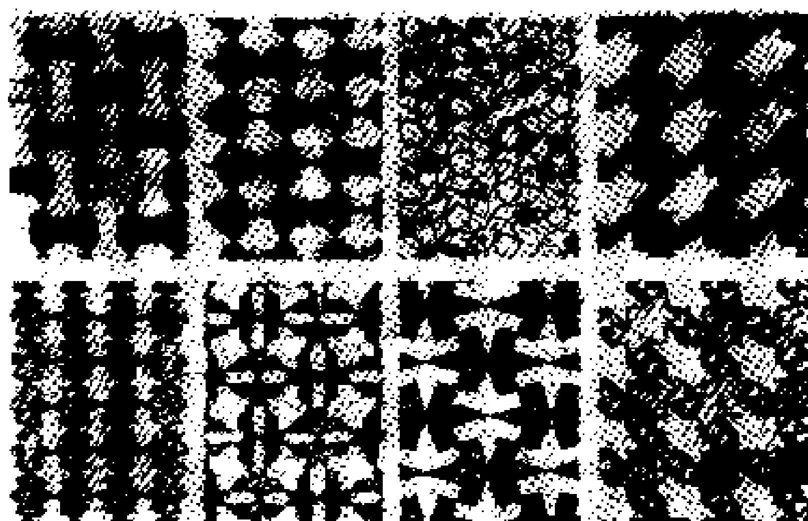


图 2

来自阿尔亨布拉宫的八幅铺砌图案.这些素描是荷兰画家M·C·埃歇尔在1936年参观阿尔亨布拉宫时绘制的.它们对这位画家的后期创作产生了深远的影响.

1. 什么是铺砌?从数学角度来看,它是一族得以既无重复,亦无空隙地铺满(即重叠或空隙部分面积为0)一个平面区域的闭集族 $T=\{T_1,T_2,\cdots\}$,这里的 T_i 称为基本铺砌单元.可是,这个定义有点过于一般化,因此本文只限于讨论 $T_i(i=1,2,\cdots)$ 是拓扑圆盘的情况.也就是说,每个 T_i 都可以从一个圆盘通过连续变形而得出.特别,每个基本铺砌单元要求是连通与单连通的,因此不允许分离成两部分以上区域或者区域中有洞的情况.最常见的铺砌图案是由为数很有限的不同形状的基本单元拼合而成.这种情况可以方便地描述如下.令 $S=\{P_1,P_2,\cdots,P_k\}$ 是一(有限)族闭集,这里 T 中的每一个基本铺砌单元 T_i 都合同于 P_i 中的一个$(i=1,\cdots,k)$.则 S 就叫做 T 的原始砌块的集合,S 确认了铺砌法 T.如果 S 恰含有 k 个不同的集合(其中任意两个集合都不合同),而且所有的集合都在 T 中被用上了,这时 T 就称为是 k 块合成的.

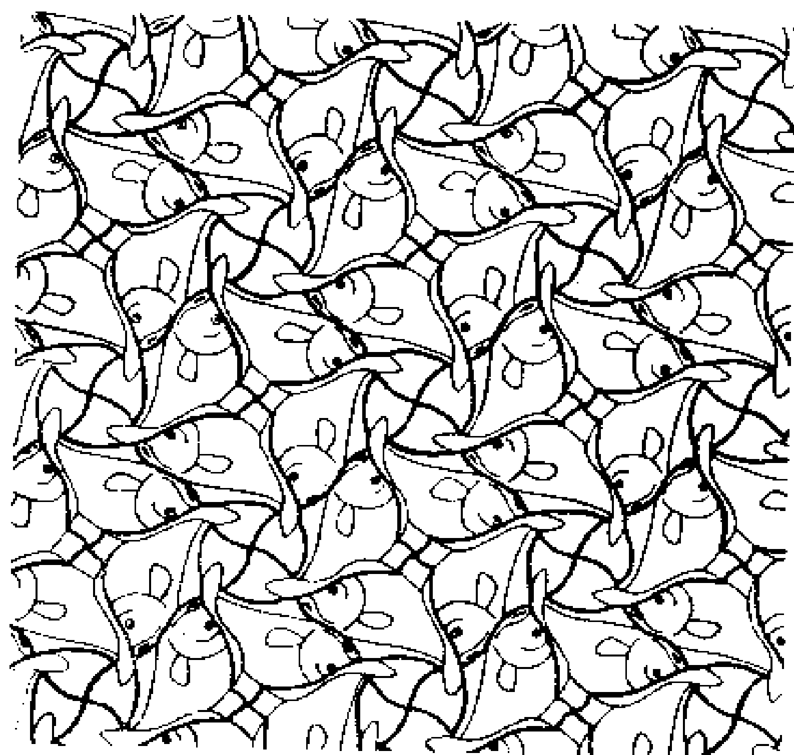


图 3

埃歇尔较少为人知晓的一种铺砌图案，在其创作中有许多动物与无机物的拼镶图案.



图 4

著名天文学家约翰尼斯·开普勒，他对平面铺砌问题的先驱性研究被世人遗忘几达三世纪之久.

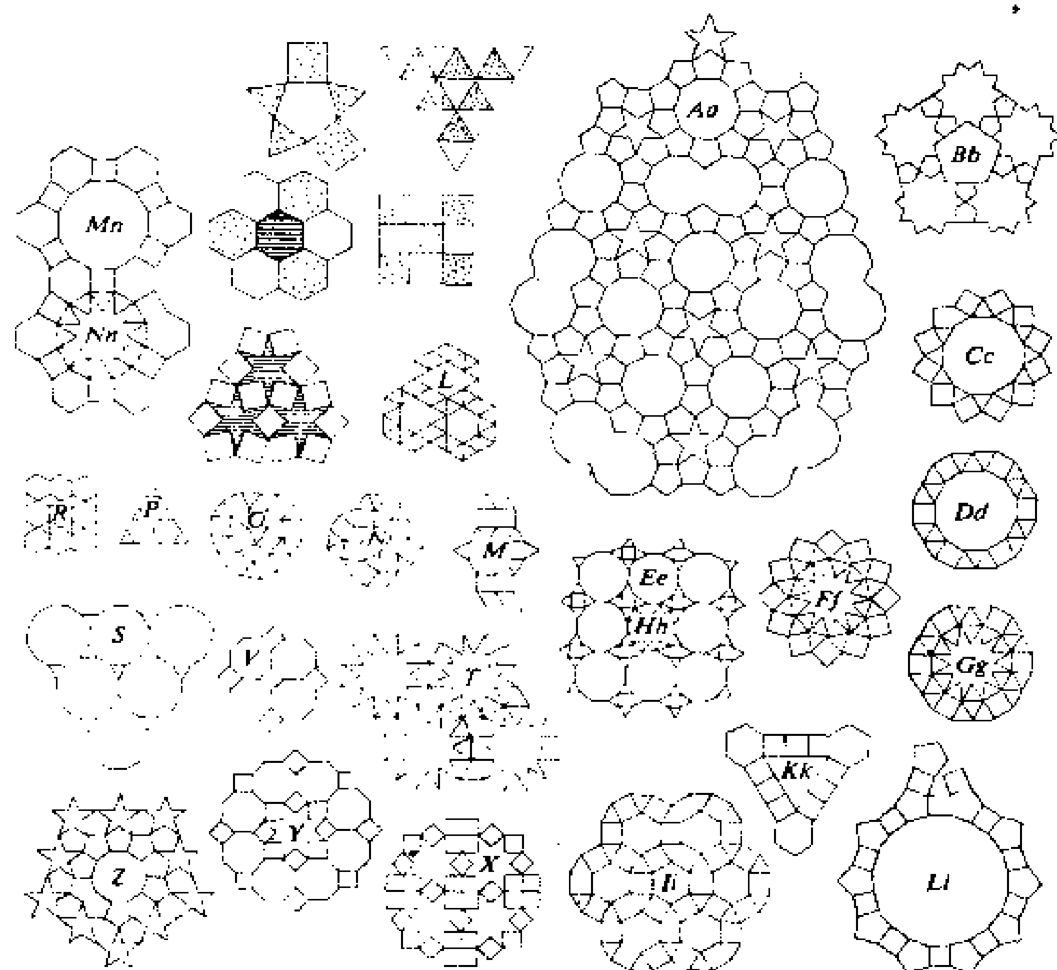


图 5

1619 年问世的开普勒著作《宇宙和声》中，首次发表了铺砌问题的数学探讨. 这本书中的插图(复制于此)给出了开普勒所曾研究过的，由规则的与星状的多边形所作成的一些铺砌图案.

$k=1$ 时，铺砌图案称为单块合成的. 此种图案极为常见. 其实例可由正则铺砌(见图 6)来提供，自久远以来，早已为人们所熟知. 均齐铺砌(见图 7)开普勒是知道的(见图 5)，也可能就是他发现的，是 2 块或 3 块合成的图案. 其他一些 k 块合成的铺砌图在 k 值较小时也是已知的，例如图 5 中开普勒画的 K, T, X 与 Aa 等图幅.

人们大概会认为，单块合成铺砌图(其中所有的铺块都是同一种

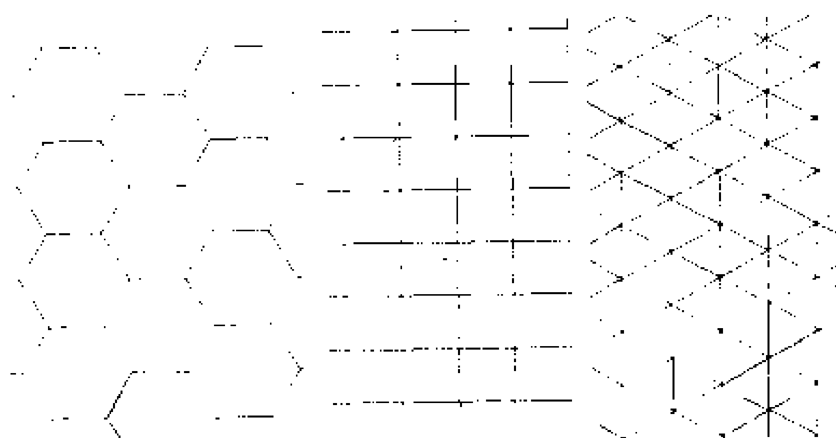


图 6

三种规则铺砌图.它们久已为人们所熟知.

形状)在数学上看来很肤浅,因而不值得加以研究.但这远非事实.图21、23、28与29给出了一些实例,它们仅仅是众多可能性中的极少一部分而已.为了使读者理解问题的困难所在,我们在图8中给出了一些原始砌块.其中的一些砌块现在还不能确切判断它们究竟能不能作为单块合成的组件.对图中给出的任何一个砌块要回答上述问题都不容易.这类问题在马丁·加德纳的《科学美国人》专栏文章(见该刊1975年8月号)中有所论述.基本问题如下:

> **问题1** 有没有一个明确定义了的过程或算法,使人们不必依赖反复试探,而直接判断一个给定的原始砌块可以作为单块合成铺砌图案的组件?

这是一个非常困难的问题.如想详细讨论它的各种细枝末节,势必进入数理逻辑的堂奥.虽然在这里作介绍并不合适,我们在第5节中将简略地指出这一问题与寻找非周期砌块之间的联系.在判别一个给定的原始砌块能否铺砌平面的、为数甚少的已知条件中,值得提出的是所谓康威准则(见图9).这一准则令人惊讶地具有许多应用,但它仅仅是铺砌存在的充分条件而非必要条件.因而,它并不能为问题1提供一个答案.

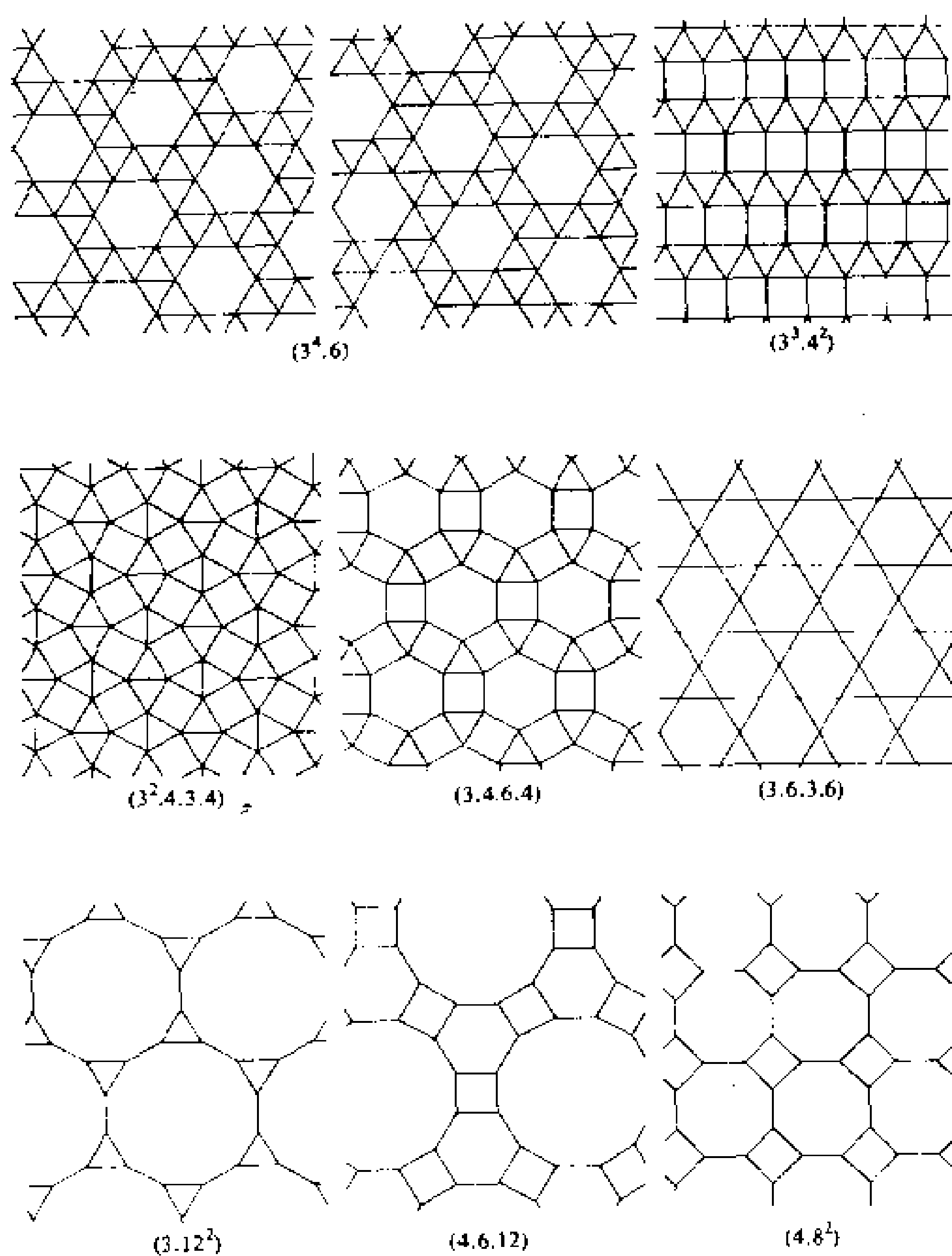


图 7

八种均齐铺砌图,它们都不是规则的. 在每幅图中,铺块都是规则的正多边形,并存在有从一个指定顶点映射到其他顶点的对称要素. 如图所示,记为 $3^4.6$ 的一种铺砌图具有两种镜面反射形式. 一般认为这组铺砌法是 Kepler 在 17 世纪早期所发现的.

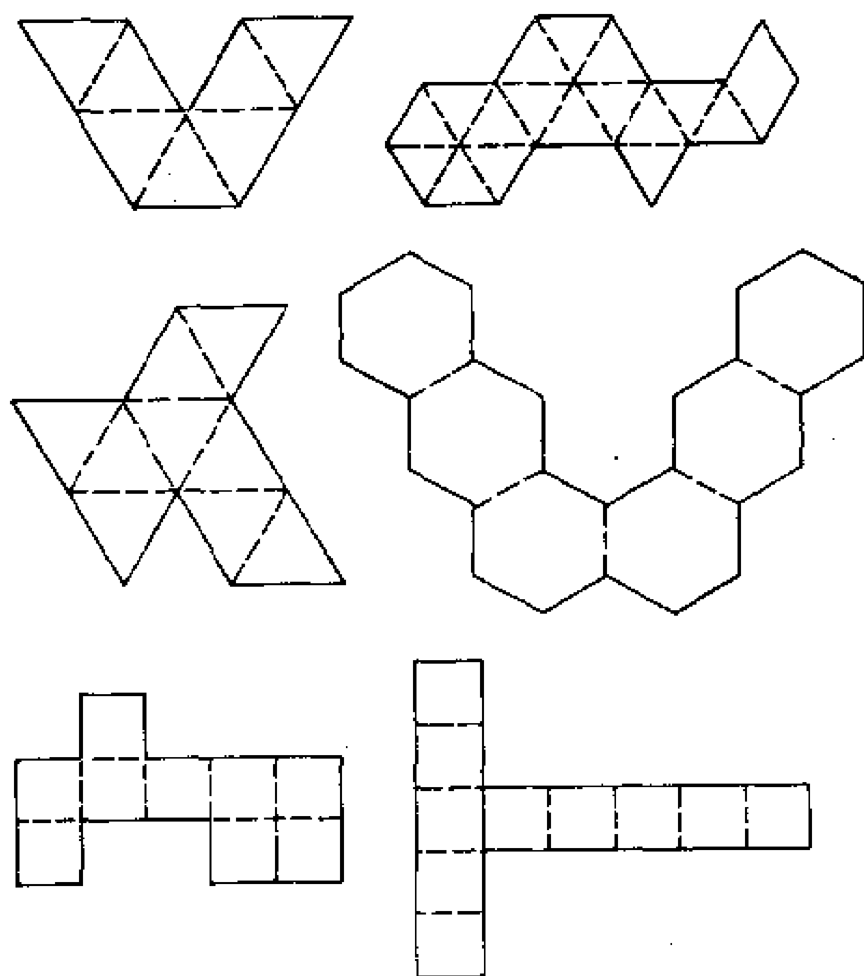


图 8

这些原始砌块，是否可以用作单块合成铺砌图的组件?要回答这个问题决非易事.

问题的深度以及我们对它的无知程度可用下列事实说明：我们甚至不知道哪些形状的多边形可以作为单块合成的基本组件. 任一三角形或任一四边形都能容许这种铺砌，还有三族六边形也是如此(见图 10).(一族多边形被定义为一组多边形的集合，它们能满足边长与角度之间的特殊关系.)但是五边形的族数则依然未能肯定. 目前，已知 13 族五边形可用于平面铺砌(图 11)，但这张清单是否完备则仍然不得而知. 马丁·加德纳对这一问题再次作出重大贡献，他在

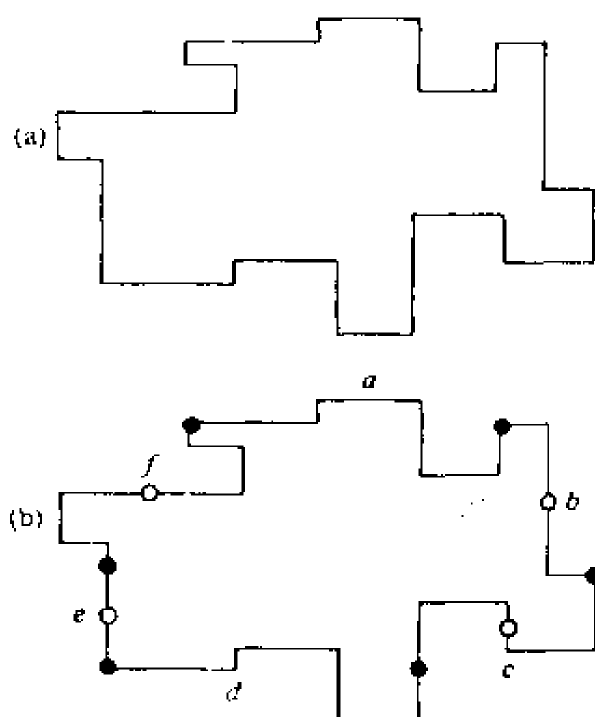


图 9

能否利用图(a)的砌块来铺砌平面? 康威准则断言这是可能的. 该准则肯定了一个原始砌块可用于单块合成铺砌图,如果其边界可以分为六个部分(由图(b)中的黑圈来表示),六部分之间有下述关系:a 与 d 是相互平移的结果,另外四部分 b、c、e、f 都有对称中心(由一个空圈来表示),在铺砌图中,六条相邻边中的四条边可由给定铺砌对空圈作 180°旋转而得. 而另外两部分则可通过平移得出. 这一准则可应用于任何铺砌,而不仅仅限于此处所提到的多边形砌块.

《科学美国人》杂志(1975 年 7 月号)上公开发表了两张表,当时认为它们已经是完备的五边形清单了. 这两张表分别由 K · 来因哈脱(K. Reinhardt)于 1918 年, R · B · 克希纳于 1968 年相继发现. 一些读者指出表格是不完备的,新出现的铺砌块包含在图 11 中. 这就导致以下的第二个问题.

问题 2 是否存在图 11 中尚未包括进去的凸五边形族,它们也可以作为单块合成铺砌图的基本组件.

最近 D · C · 亨特 (D. C. Hunt) 与 M · D · 希尔锡洪 (M. D.

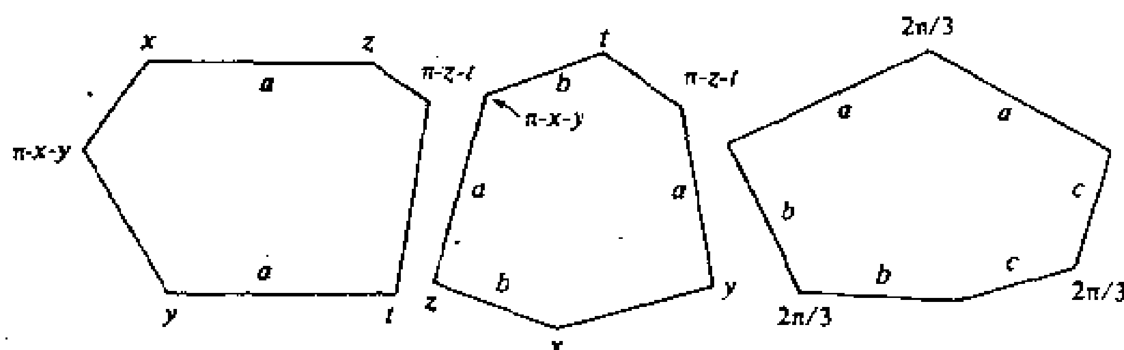


图 10

可以通过单块重复来铺砌平面的三类六边形.实际上每个图形还是等面式的铺块.这些类型于1918年为K·来因哈脱所发现.图上分别表明了各内角(记为x,y,z,t)与各边(记为a,b,c)之间的关系.

Hirschhorn)声称已称证明了多丽丝·沙特斯奈德在1978年编纂的一张可以铺满全平面的等边五角形表格是完备的.然而,他们的证明却还没有印出来.

2. 即使一个原始砌块可以用作平面上的单块合成组件,也并不存在一种先验的方法来判断它究竟可以用多少种不同方法来铺砌.如果只存在唯一的铺砌方法,则我们称此种铺块为单一同态的.最常见的单一同态铺砌是正六边形——唯一可能的办法是像图6中所示的规则铺砌.另一方面,正方形却不是单一同态的.事实上,通过"滑动"一排正方形,可以得到无穷多种不同铺砌法.在图12中我们举出了一些较少为人知晓的单一同态铺砌.同时,在考虑它们时,也产生了一个重要的定义问题.譬如说,图12中,铺砌(e)同(f)似乎都有着两种铺砌方式(见图13),但两者却是互为镜面映像的.我们是否应把它们看作不一样?对此,最好的回答基本上属于一种观点性质,但根据种种理由,我们还是不要把镜面映像视为不同为好,因此我们认为铺砌(e)、(f)仍然是单一同态的.

是否存在双相同态的铺块?也就是说,不多不少,恰是两种方

图 11

目前已知的十三类五边形，它们可用作平面的单块合成组件．前五类还是等面式的铺块．此图取自沙特斯奈德的论文（见《数学杂志》51卷（1978年），第29—44页）．该文还给出了进一步的细节与铺砌实例．



图 12

九种单一同态铺块.

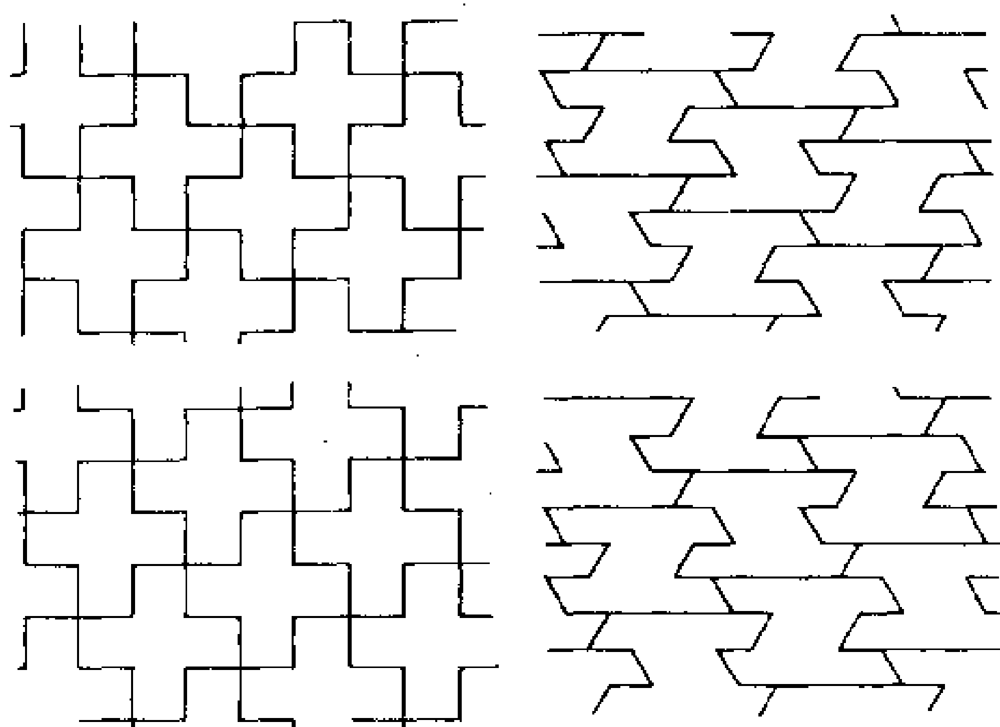


图 13

图 12 的两种铺块(e)与(f)可以各有两种铺砌法.但它们互为镜面映像,因而铺块仍被看作单一同态的.

式?其答案是,此种铺砌确实存在,在图 14 中给出了实例.一种三相同态铺砌及其相应的三种铺砌法例示于图 15.这个例子在某种意义上说是唯一的,因为所有其他已知为三相同态的铺块,同它只有略微的差别.双相同态与三相同态铺块的发现是不容易的,于是我们提出下面的两个问题.

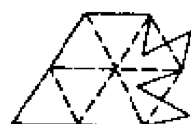
问题 3 找出另外的、实质不同的双相同态与三相同态铺块.

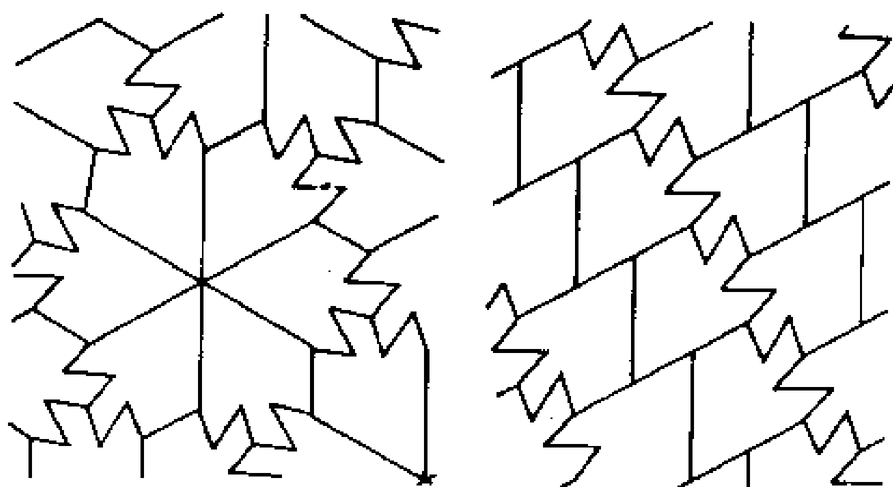
问题 4 对任意有限值 $r \geq 4$,r 相同态的铺块是否存在?

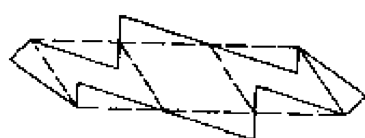
下面是一个更常有技术性的问题,但它也有一定理论价值.

问题 5 是否存在着一种铺块,它只能具有可数无穷种铺砌法?









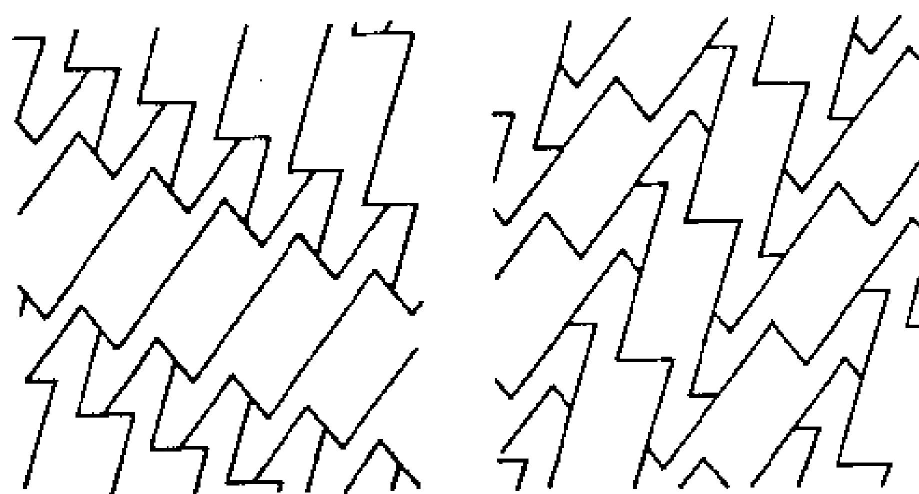


图 14

两种双相同态铺块及其相应的铺砌方法.

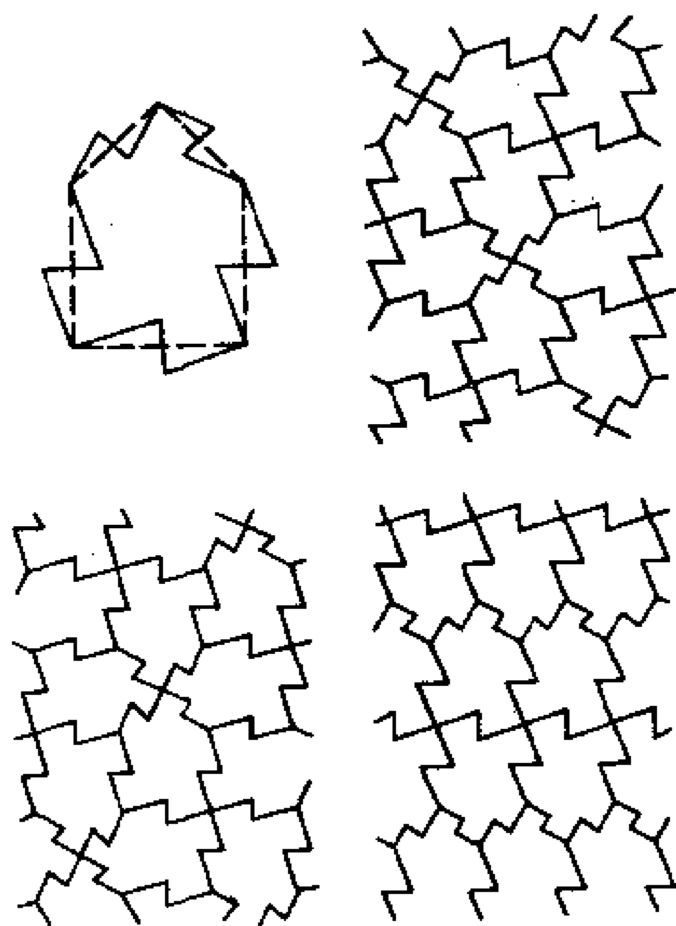


图 15

一种三相同态铺块及其相应铺砌法.

由此可见，问题 4 与 5 是在质询能否架设一个桥梁，以跨越介于 $r=3$ 与一个不可数无穷大之间的巨大缺口.

如果我们试图把以上这些概念推及于 k 块合成铺砌（在铺砌平面时要使用 k 种形状的铺块），那就需要一些新的考虑. 让我们来研究 k 块合成、r 相同态的铺块（$k\geqslant 2, r\geqslant 1$）. 它们的存在曾于 1977 年为 H. 哈包思（H. Harborth）所阐明. 他指出，适当选取一个菱形与由一些菱形组成的砌块（见图 16），即可回答 $k=2$ 的问题，从而也就解决了 $k\geqslant 2$ 的问题.（这只要把上述任一种铺块分割成若干小块，使这些小块重新集合起来时只能恢复为原先的铺块.）然而，图 16 的作法并不能完全解决问题. 仅当我们坚持这一点：在所有的铺砌图案中，

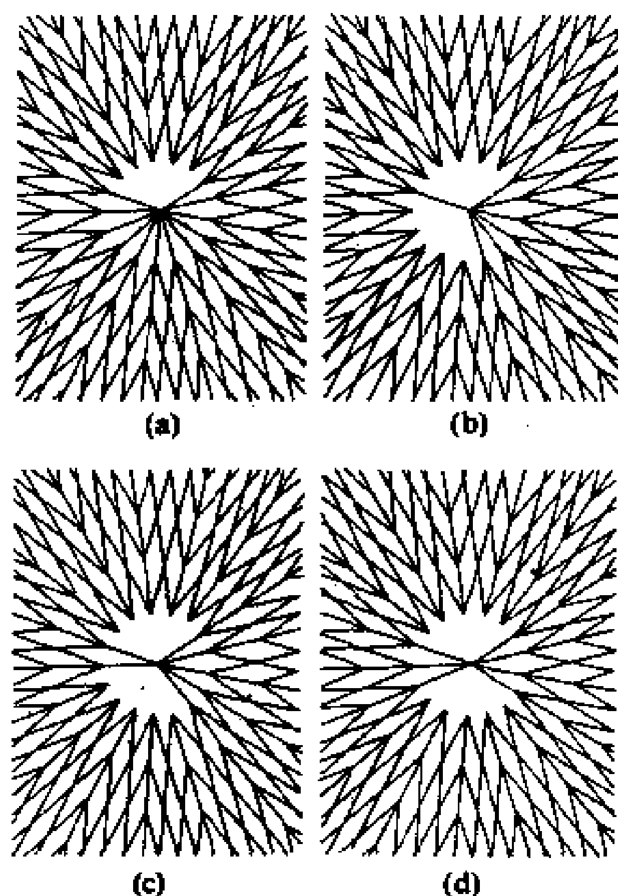


图 16

不同铺砌法恰有 r 种的一对原始铺块的哈包思构形法(假定每个铺砌图案中都出现两种铺块). 在此图中, $r=4$. 一个铺块是内角分别为 $2\pi/p$ 与 $(p-2)\pi/p$ 的菱形 $(p=6r-7)$, 另一个铺块则由 $2r-2$ 个菱形聚合而成.

每个原始砌块的拷贝都必须用上时，问题才算完全解决了. 如果没有这一条件，那么菱形可以滑动，从而产生不可数无穷多种铺砌方法. 我们于是再提出下列问题.

问题 6 对每一个 $k\geq 2$ 与 $r\geq 1$ 的值，是否存在 k 个原始铺砌的集合 S，使得从 S 可以得出 r 种铺砌方法，即便 S 中所有原始铺块的拷贝并不需要完全用上？

当 $k=r=2$ 时，问题 6 已告解决，其实例见图 17.

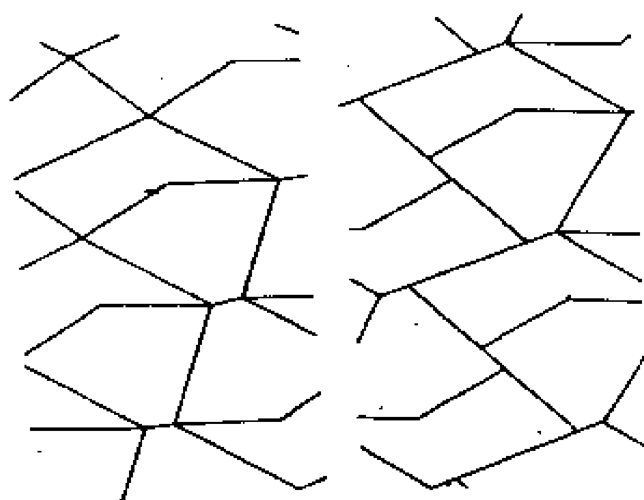


图 17

利用同样的两个五边形铺块所作成之两块合成铺砌图. 仅有这两种可能性，因而一对五边形构成一个双相同态集合. 请注意这两种五边形铺块的任一个，若不与另一个结合，都不能独立完成铺砌平面的任务.

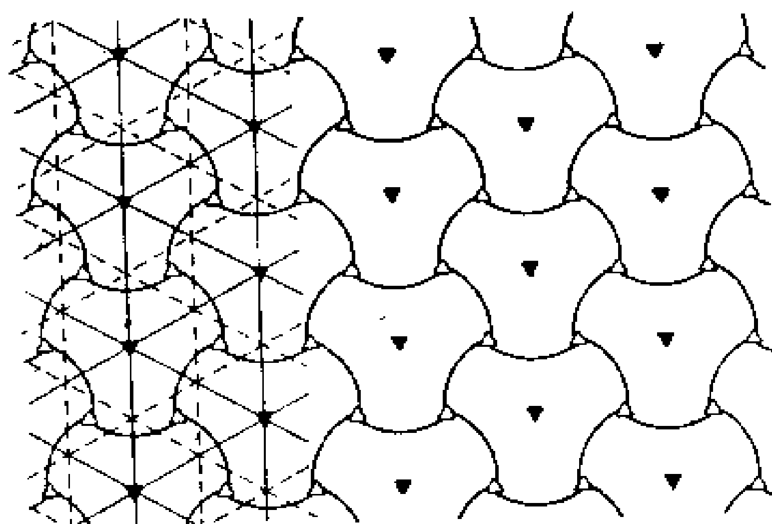


图 18

一个铺砌图案及其对称要素(除平移外，都已标明在图的左半部分). 三角形标明 3 次旋转对称的中心；实线表示反射直线；虚线则代表滑动反射. 连结两个实心三角形的向量则是一个平移对称.

3. 众所周知,大卫·希尔伯特(D. Hilbert)在1900年提出了一系列问题,对数学的发展曾起过很大作用. 这些问题中的一个——第18问题与铺砌有关. 下面是该问题的简略介绍.

为了说明问题,我们必须使用"等面"这个单词. 它可定义如下. 每个铺砌 T 都具有一个对称要素所构成的群 $S(T)$,也就是说,平面的刚体运动可以把 T 映射到其本身(例如,可以参看图18,该处已标志了几种不同的对称要素). 福莱(Fourrey)于1907年提出了研究对称的一种非正式办法(数学上的严密性不无保留). 他建议把铺砌图案用透明纸影描下来,于是,铺砌图的每一种对称即相当于透明纸的一种运动(也包括把它翻转过来的可能性)以使得影描下来的图形再次与铺砌图完全一致. 如果给定 T 的任意两个铺块 T_1, T_2,存在着 T

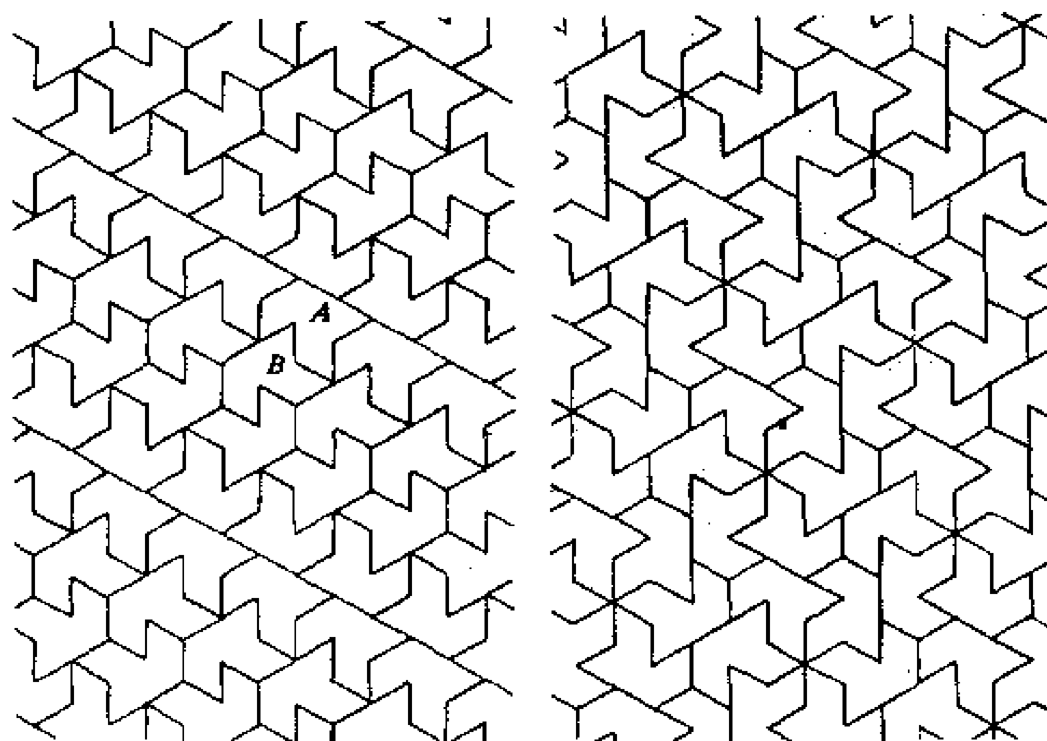


图 19

利用同一种原始砌块的两个单块合成铺砌图案. 右方为等面的铺砌,因为任意给出两个铺块时,存在把其中的一个映射为另一个的对称要素. 但左方不是等面铺砌. 因为,没有一种对称要素可使铺块 A 映射成铺块 B.

中的一个对称要素能使 T_i 映照成 T_j，则称铺砌 T 是等面的. 任意一种等面的铺砌必然是单块合成的——这两个概念之间的区别在图 19 中有所说明.

希尔伯特第 18 问题可叙述如下. 是否存在着一种原始砌块，它可以作成单块合成铺砌图，但却不能接受等面铺砌图. 事实上，该问题原先提出的背景是三维空间. 有理由假定，在平面的情况下，可以毫不吃力地得到反面回答，希尔伯特相信如此. 可是他错了！1935 年，希许(H. Heesch)发现了一个反例(见图 20). 有一种铺块可作出无限多铺砌图，但却没有一个是等面的. 其后，又发现了其他实例.

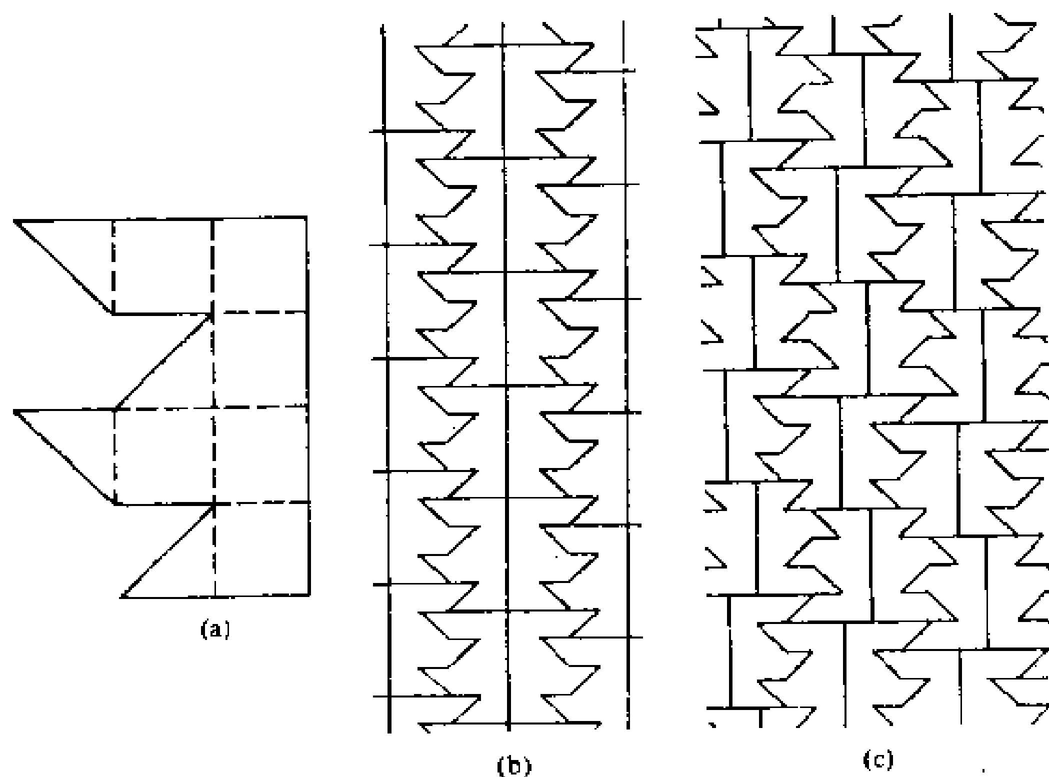


图 20

这里给出的希许氏铺块，是人们发现的第一个实例，它可以作成单块合成铺砌图而不是等面铺砌图. 现在已发现了许多其他例子.——譬如说，图 11 中的某些五边形即有此种性质. 我们在图(a)中标明该铺块怎样从单位正方形与半个单位正方形作出，然后在(b)，(c)中作出了(无限多种)可能的铺砌图.

图 11 中，有一些凸五边形即具有此种性质.

初看起来，似乎单块合成铺砌与等面铺砌的差别甚小，但这远不是事实. 例如，假定我们把注意力限于等面铺砌的情况，则问题 1，2，4，6 可以解出. 特别地，对问题 2 来说，H·希许与金兹莱(O·Kienzle)在 1963 年已作出一张完整的清单，其中的多边形可用于等面铺砌. 之所以有此可能，是因为所有的等面铺砌图均能描述. 可以

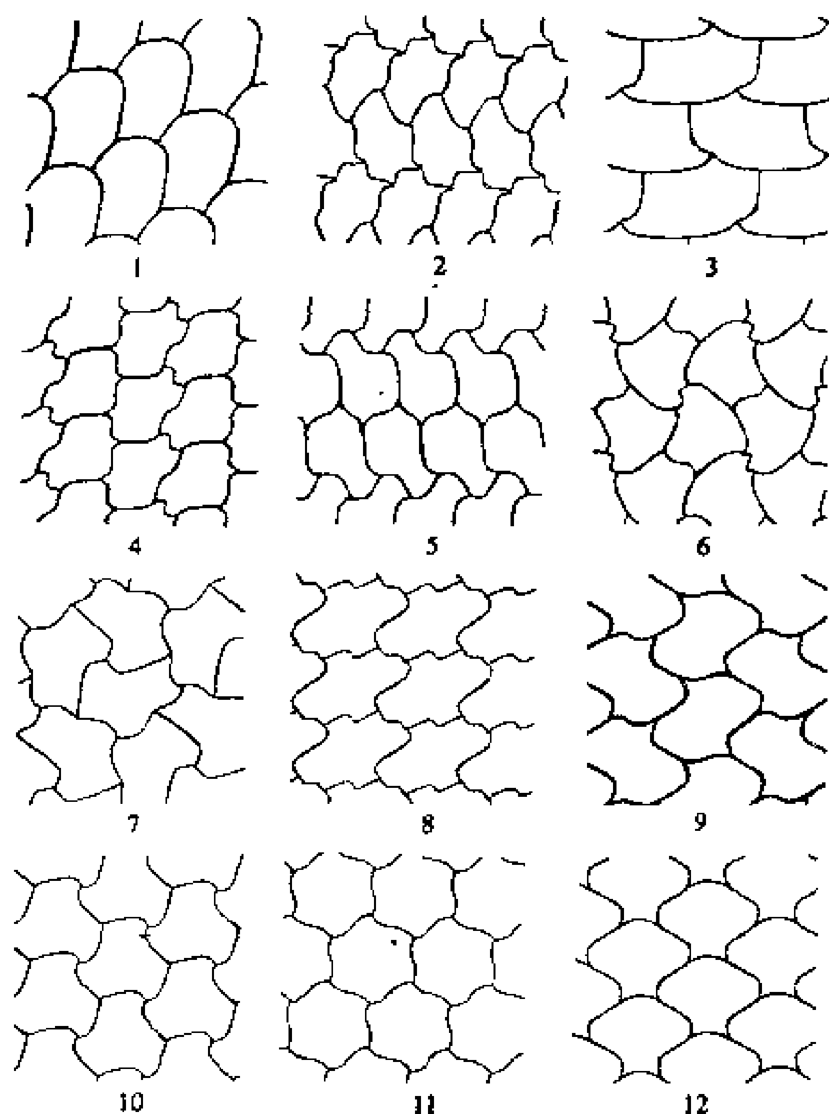


图 21

等面铺砌的十二个例子.

证明它共有 81 种类型，其中的 47 种可以用多边形加以实现. 在图 23 中给出了若干实例. 至于如何具体分类则因技术细节过于复杂，这里不拟详述. 它只是在最近的一篇综述性文章中才得以圆满解决(这一事实也许可以解释论述分类问题的一些旧论文中何以充斥着许多错误与含糊不清之处).

4. 有关铺砌的、最值得注意的定理之一是所谓*扩展定理*. 它的一个特例叫做王氏定理(以发现者的姓氏命名)，早已为人们所知晓，但一般情形仅在最近才得以证明，而其证法尚未公开发表. 设 S 是事先给定的一个原始砌块的(有限)集合，每个铺块都是一个拓扑意义上的圆盘，并假定，不论 R 有多大，我们总是能够利用 S 中的铺块来铺完半径为 R 的圆盘 D_R，于是扩展定理断言 S 也可用于铺满平面. 这里所说的"铺完"一词需要加以澄清. 其意思指：有可能作出一个铺块的集合(或一块补钉❶)，以使得它们不相重叠，然而却能完全

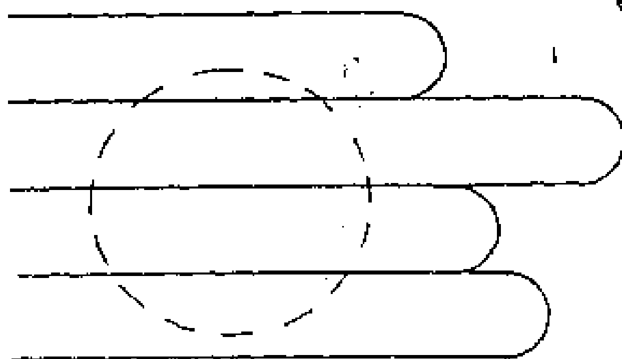


图 22

这里的每一个铺块是末端为半圆的半无限长带条. 这样的铺块显然能铺完任意大的圆盘，然而却不能用来铺满全平面. 本例表明要使扩展定理成立，必须假定铺块❷是有界的.

---

❶ 译者注：此处使用"补钉"一词，看来很奇怪，其实是指铺砌图案的形状极像补钉而言，请看图 23.

❷ 译者注：由此足以看出，作者使用"铺块"与"原始砌块"非常随便，两者并无严格区别，纯属文风.

覆盖一个包含 D_R 的区域. 要使本定理成立，其关键是原始砌块必需有界，且其个数有限. 例如，在图 22 中我们举出了一个原始砌块，利用它的一些拷贝可以铺完任意大的圆盘，可是却不能铺满全平面. 当然，这一情况是用不上扩展定理的，因为原始砌块是无界的.

本定理的有些推论令人惊奇. 例如，定理蕴含着如下推论：如果我们能利用 S 的铺块铺完四分之一平面，则我们也将能铺满全平面. 初看起来，这是再明白不过的：如果我们可以毫无限制地，作出越来越大的铺块补钉，那么最后当然可以铺满全平面. 但如果意识到当 R 增大时，也许有必要不断重新安排补钉中的各个铺块，而有可能覆盖圆盘的所有补钉都不是最后铺砌图的一部分的话，当初认为明显的结论也就变得模糊不清了.

与此有关的是所谓希许问题. 已知以下事实：存在着原始砌块 P，它不能用来铺满全平面. 然而 P 可以被它的一些拷贝完全包围

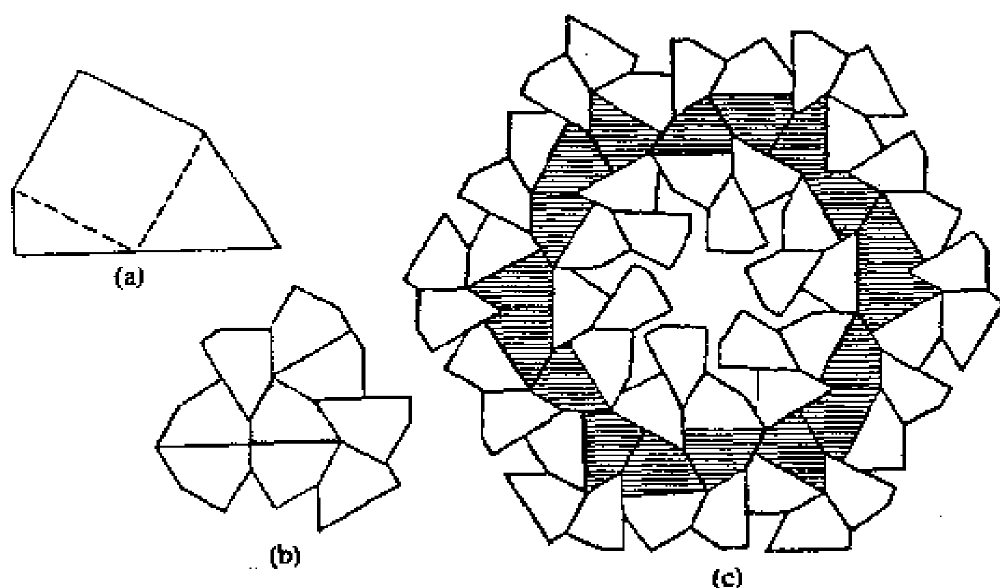


图 23

图(a)的希许铺块具有以下值得注意的性质，它可以被其自身的拷贝完全围绕，但却无法铺满平面. 即使这样，尚可作出更大的补钉. 如图(c)所示. 铺块的形状由图(a)给出，它是由一个正方形、一个等边三角形与一个等边三角形的一半所合成的，见图中之虚线.

（见图 23），所谓完全包围 P 是指，我们能够作出由铺块所构成的一个环来围绕 P（不留下空隙），以使得 P 的每一点与 R 外面的、平面上的任一点的距离大于某个固定的正数. 所谓 P 被完全围绕两次，则意味着：P 被一个环 R 围绕，而 R 又被第二个环 R' 围绕.

问题 7 是否存在一种原始砌块 P，它不能用来铺砌全平面，但却能被 P 的拷贝完全围绕两次？

更一般地讲，当我们说一个铺块 P 可以被围绕 r 次，其意义当然是很清楚了. 如果 r 是此类整数中最大的一个，则称做铺块 P 的希许数.

问题 8 是否存在希许数 $r=3,4,5\cdots\cdots$ 的铺块？

5. 如果 T 的对称要素中存在两个互不平行的平移，则铺砌 T 称作是周期性的. 周期铺砌的实例在图 13、15 与 18 中均可看到. 以上每一场合，人们都可以把铺砌图案看作由铺砌组成的一片补钉，按格子配列在平面上反复平移而得. 原始砌块的一个集合 S 称作非周期性的，如果它能作成平面的一种铺砌图，然而却不是周期性的. 迄今为止未发现过非周期性的集合，第一个例子由 R·伯杰（R. Berger）于 1966 年发现，其后别的例子又被罗宾孙（Robinson），彭罗斯（Penrose）[1]，阿曼（Ammann）与别人相继发现. 图 24 给出了由 R·彭罗斯所发现的非周期性铺砌中的第一个集合所作成之平面铺砌图案.

对于彭罗斯非周期性铺块中"风筝"、"标枪"等构形的趣味横溢的故事，读者们不妨参阅《科学美国人》1977 年 1 月号上马丁·加德纳的文章. 这些铺块具有一些值得注意的、出乎意料的性质，并非全都已经作了充分解释. 1977 年以来，又有几组新的非周期性铺块被罗伯特·阿曼（Robert Ammann）所发现，承其美意，慨然允诺我们在图 25 中复制一些铺块. 于此谨向他致谢.

非周期性铺块的题目实在太大，此处无法详加讨论. 本问题的历

---

[1] 译者注：罗杰·彭罗斯（Roger Penrose）当代著名代数学家，英国剑桥大学教授，广义逆矩阵理论的奠基人.



(a)　　　　　　　　(b)

(c)

图　24

由罗杰·彭罗斯所发现的第一组非周期性铺块. 图(a)给出了这些原始砌块,图(b)则用数目字来表示突出与嵌入的部位,从而指出了一种"匹配条件",0,1,2 必须与 $\bar{0},\bar{1},\bar{2}$ 分别配合. 图(c)则显示了具体的铺砌图案.

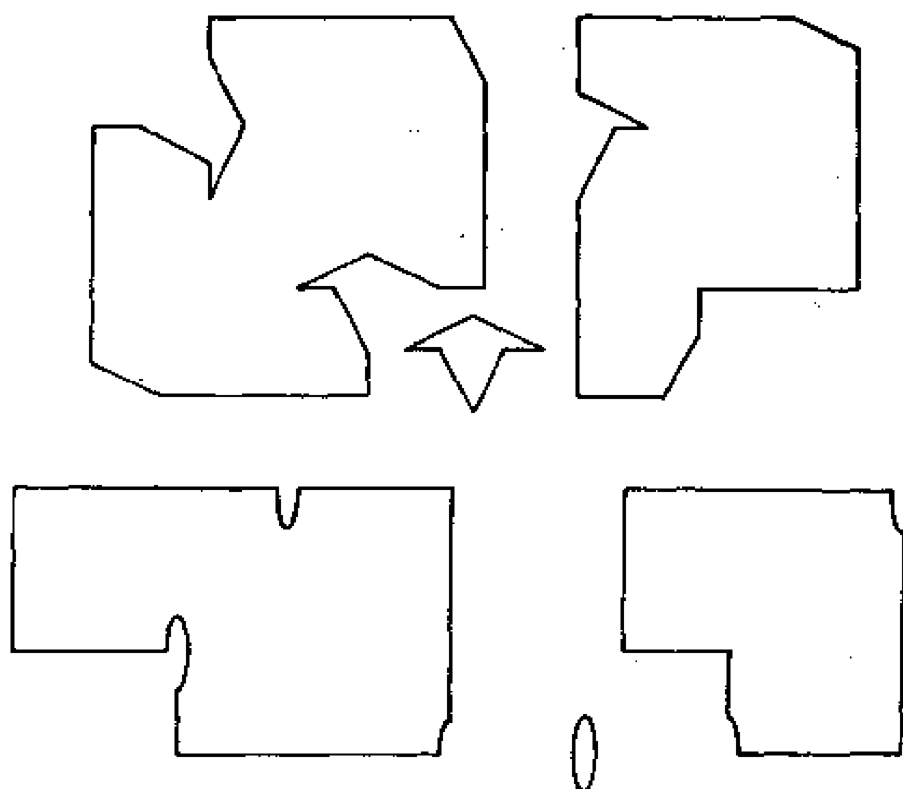


图 25

最近由 R·阿曼发现的两组非周期性铺块. 每组共有三块,其中有一块较小,名为"钥匙块",它可以榫合较大铺块上的缺口,从而对两块较大组件的相互贴近方式作了限制.

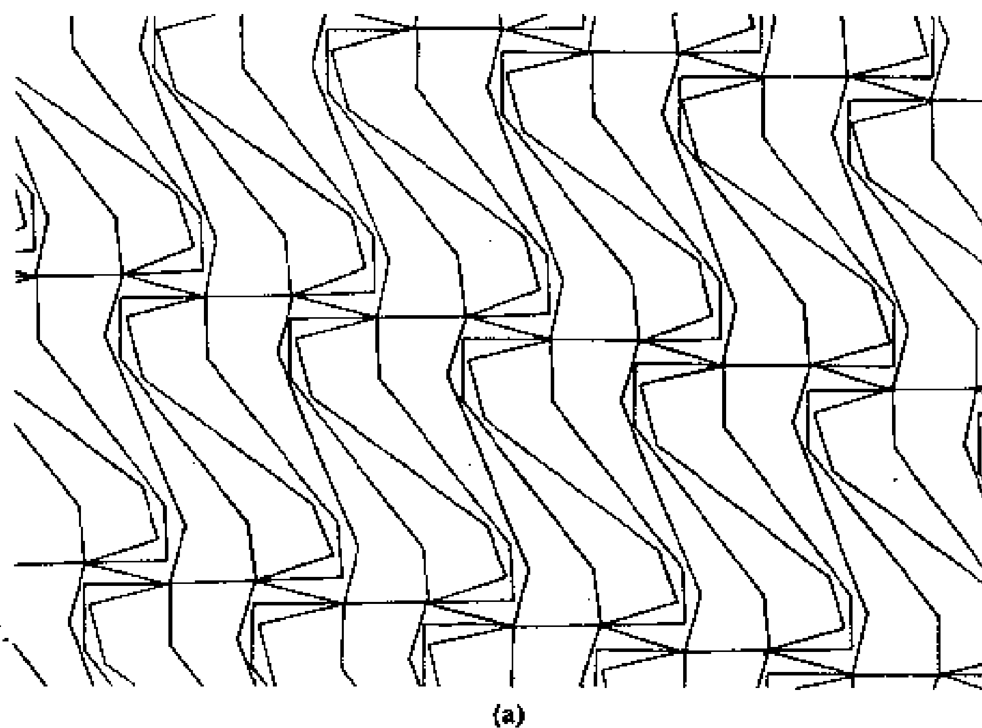
史以及它同数理逻辑的联系在整个铺砌问题史中是一个动人的篇章,而且这一章至今仍未完全写好. 此处我们只能略提一些事实,而无法做得更多. 尽管如此,如果毫不提及这个领域中的主要突出问题,企图对铺砌性质作一综述的任何文章都将是不完整的.

问题 9 是否存在单独一块原始砌块 P,以构成一个非周期性铺块组?

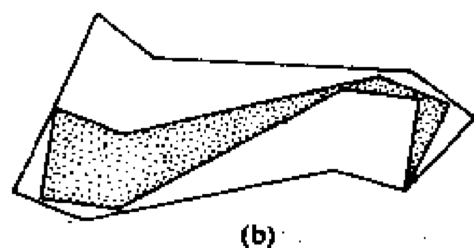
换言之,P 可以作出平面上许多由单块构成的铺砌图案,然而其中没有一个是周期性的. 令人惊讶地,可以证明这一问题与问题 1 有关,最低限度,王浩(H. Wang)[1]业已证明非周期性与铺砌问题有联

[1] 译者注:著名数理逻辑学家王浩.

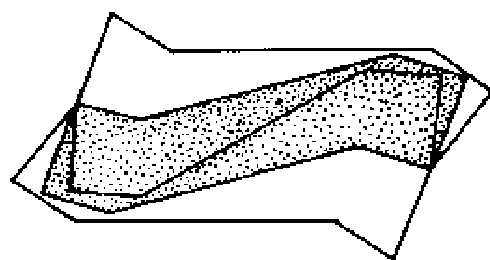
系，即决定一个算法以判明给定的一组原始砌块能否铺砌平面. 王氏的结果仅能应用于非常特殊的“着色边的正方形铺块”，我们一点不



(a)



(b)



(c)

图 26

伏特堡氏铺块，它具有一种值得注意的性质：该铺块的两个拷贝能把第三个拷贝全部围住(图(b))，甚至能把两个拷贝围住(图(c)). 图(a)则是把伏特堡氏铺块用作原始构形时的周期性铺砌图.

知道对于一般形状的铺块来说，类似的考虑可以应用到何种程度. 尽管如此，看来有可能，问题 9 的肯定解答将蕴含着问题 1 的否定答复.

6. 让我们用若干珍品结束本文，它们中的第一个是所谓包围问题. 来因哈脱在 1934 年提出问题：对两个铺块来说（其中的每个铺块都与一个原始砌块 P 合同）是否有可能围住 P 的另一个拷贝. 1936

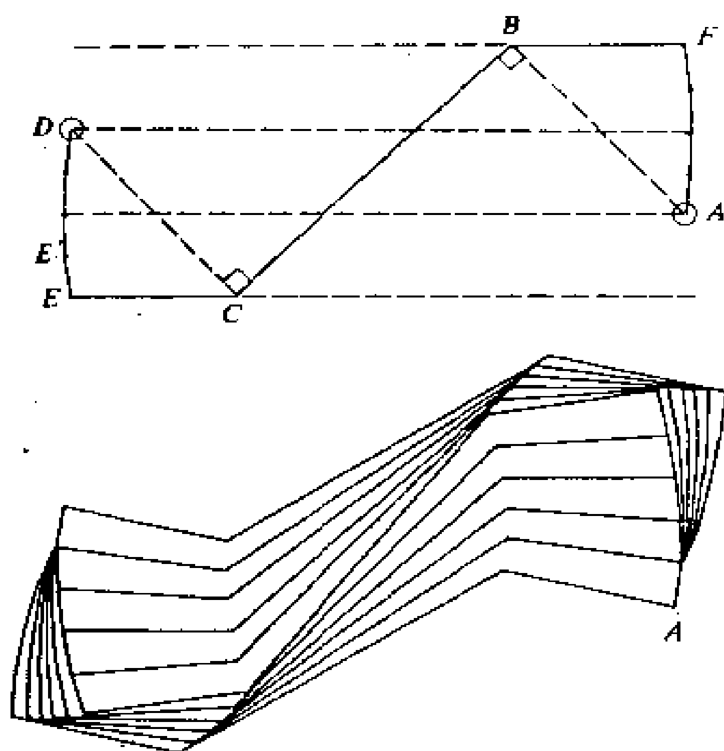


图 27

一个具有 r-块包围性质的铺块作法（图上，$R=8$）. 先作一条折线 $ABCD$，其中的 B、C 为直角顶点，点 A、B、C、D 均位于四条等距的平行线上（图上用虚线表示平行线）. DE、FA 是中心分别位于 A 及 D 的圆弧，EE' 是弧 ED 的四分之一. 用 S 表示 $AFBCE$，S' 表示将 S 围绕 A 旋转直至 E 与 E' 重合时所得的折曲线，则由 S、S' 与 EE' 所围成之铺块即具有 8 块包围性质，见图 27 的下半部分. 如果 r 值是其他数字，则仍可仿照上述之作法，只须选取 E'，使弧长 DE 为 EE' 的 $\frac{1}{2}(r+1)$ 倍即可.

年，这个问题被 H・伏特堡(H. Voderberg)解决了，他得到了图 26 这一实例，而它是一种周期性的铺砌图. 这种原始砌块具有一项奇妙性质：某些铺块不仅能包围第三个铺块，而且其他几对铺块还能包围 P 的两个拷贝. 我们可以说伏特堡氏铺块具有 2 块合围性质——一般地说，如果 P 的两个拷贝能围住一个区域(它等于 r 块没有重叠的 P 的拷贝之和)，则铺块 P 即具有 r 块合围性质. 值得注意的是尽管 r 值选得很大，具有 r 块合围性质的铺块依然存在. 图 27 表明这类铺块的一种构作法. 然而，这些铺块还是未能解决下面的(也许不太困难)问题.

问题 10 是否存在着一种具有 r 块合围性质($r \geqslant 3$)，

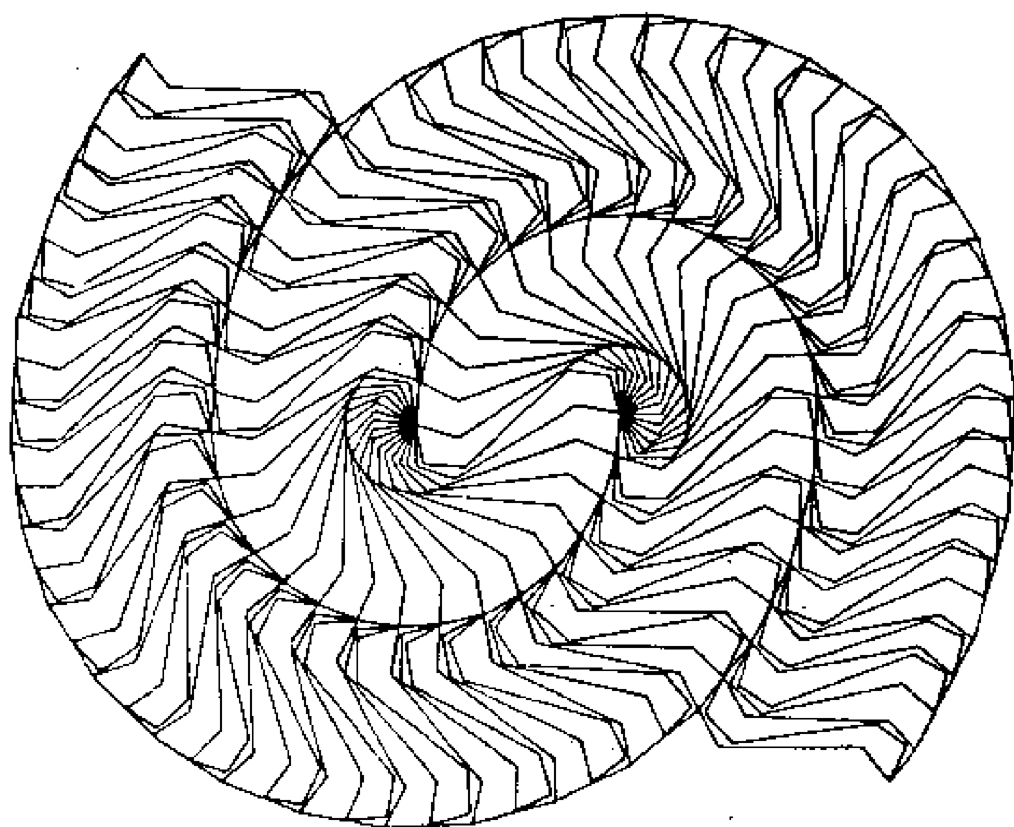


图 28

伏特堡氏的螺旋形铺砌图. 它利用了图 26 中的原始砌块.

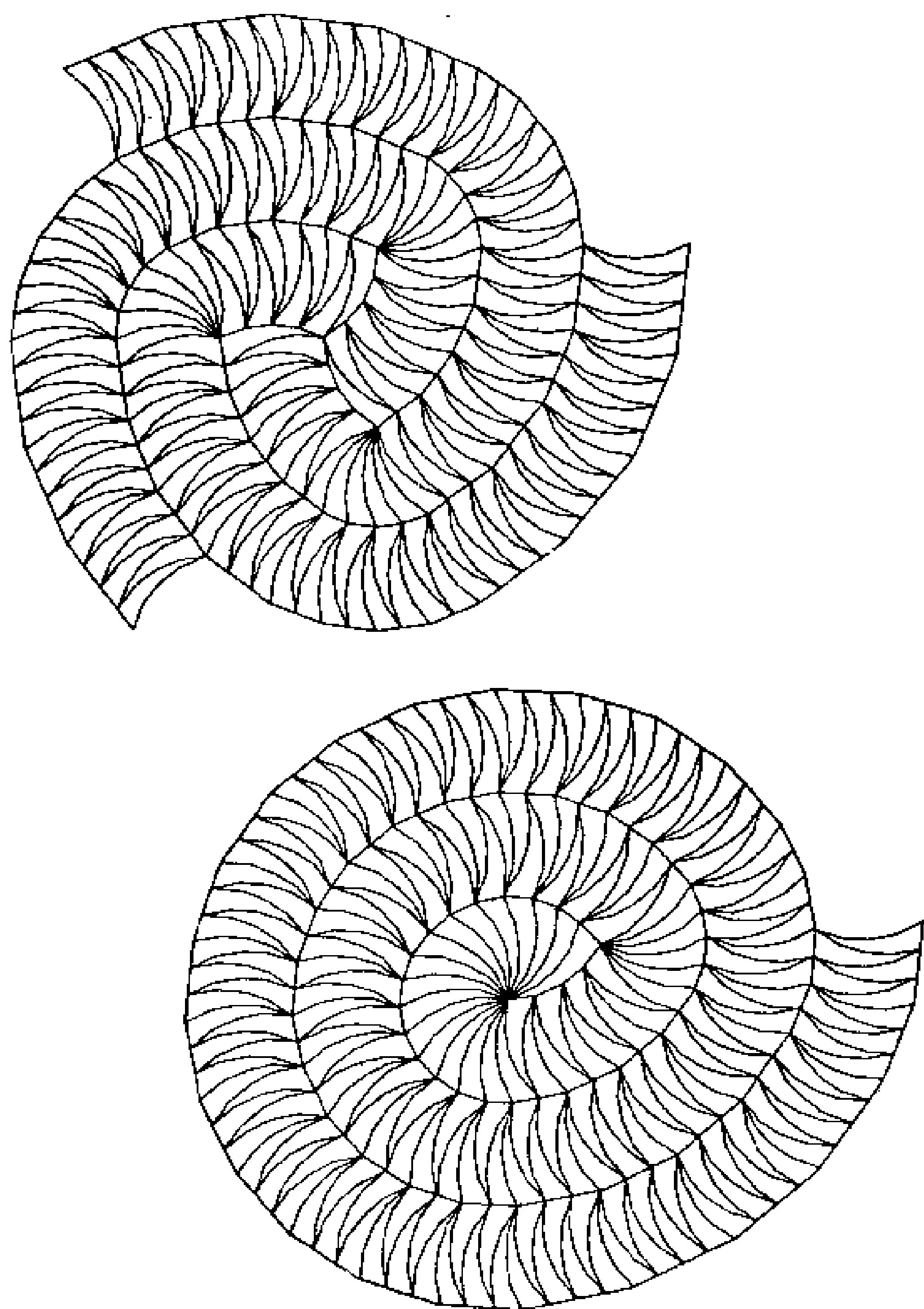


图 29

具有三个与一个旋臂的、纯由单块合成的螺旋形铺砌图，所用的原始砌块即为有名的“多才多艺者”.

并能铺满平面的铺块 P?

伏特堡在其论文中几乎是无意地注意到,他的铺块可按"螺旋形"进行铺砌(见图 28),此种外表显得颇为吸引人,从而引起了相当的注意. 1977 年 1 月,马丁·加德纳发表了伏特堡氏螺旋线的一幅图,以及迈克尔·戈德堡(Michael Goldberg)对其制作法的解说. 一旦讲明了方法,即可了解到螺旋形外表与包围性质没有什么联系. 此外,也很容易找到能够作出螺旋状铺砌图案的许多不同铺块.

虽然如此,能用戈德堡法作出的所有螺旋线都必须具有偶数个旋臂,即有一排铺块自中心向外旋出. 最近,具有奇数个旋臂的螺旋线也已发现,其实例见图29所示. 这里所用的铺块被称作多才多艺

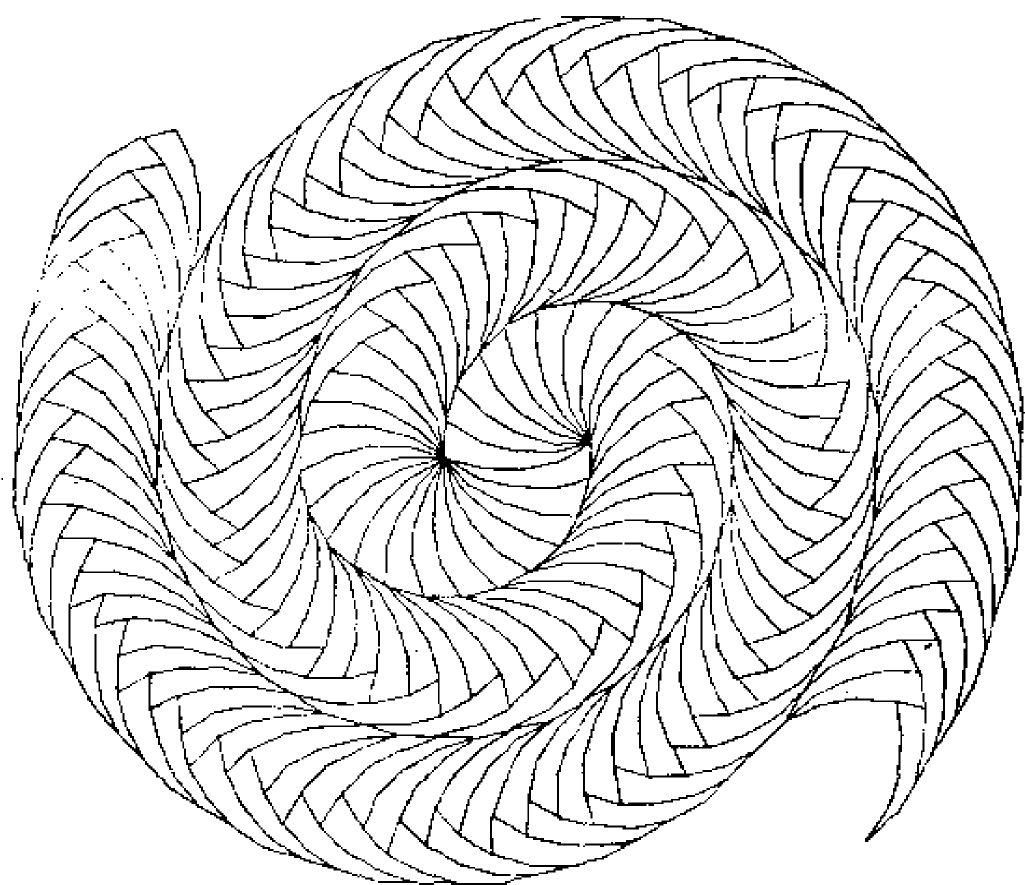


图 30

一种装饰性螺旋形铺砌图案,所用的铺块与图 29 的铺块是同一个东西.

的，因为它可以作出许多其他不寻常的铺砌图案(见图 30). 然而仍有许多悬而未决的问题，譬如说，在图 29 中我们注意到了大约有一半左右原始砌块的拷贝是其他拷贝的镜像. 是不是非此不可呢?

问题 11 是否存在一种单块合成的螺旋形铺砌图案，它具有奇数个旋臂，并且只利用原始砌块的直接拷贝(不是其镜像)?

也许我们应当提一句，螺旋形铺砌图这个课题，虽然在美学上非常吸引人，却具有很大的数学缺陷性——迄今为止，我们未能确切地说明，螺旋形铺砌图究竟是个什么东西;它是不是一个真正的数学概念，还是仅不过一个心理学的东西?最后，我们要提到下面的问题，也许它更适宜于一般讨论而不是去作数学研究:

问题 12 对螺旋形铺砌图给出一个确切的定义.

参考材料及进一步阅读之文献

除了本文引证过的马丁・加德纳的文章之外，下列图书与论文也是很有趣的. 以下根据本文的有关段落予以逐一列举.

引言 已经出版了好几部埃歇尔的作品选集，收罗最为宏富的是《埃歇尔的幻想世界》(The World of M. C. Escher, Abrams, New York, 1971)，在 B・恩斯特(B. Ernst)所写的书《埃歇尔的魔镜》(The Magic Mirror of M. C. Escher、Random House, New York, 1976)里有着非常有趣的埃歇尔传记以及他的一些铺砌作品. 在阿尔亨布拉宫中发现的铺砌图案的有关讨论可以参看 E・缪勒(E. Müller)的著作《格伦那达的阿尔亨布拉宫中的群论艺术品》(Gruppentheoretische Ornamente aus der Alhambra in Grenada、ETH dissertation, Zürich, 1944). 1619 年于林兹(Linz)出版的开普勒原著《宇宙和声》(Harmonice Mundi)，在 M・喀斯巴(M. Caspar)主编的《开普勒全集》中又作了重印(Gesammelte Werke, Band Ⅵ, Beck, München, 1940 and by Culture et Civilisation, Bruxelles, 1968). 这些文章都是用拉丁文写的. 由 M・喀

斯巴所译的一本德译本《宇宙和声》也已经出版(Weltharmonik,Oldenbourg,München,1967).

1. 可以铺满全平面的三类六边形是由K·来因哈脱在其学位论文“论平面的多边形分解”(Über die Zerlegung der Ebene in Polygone,法兰克福大学,1918(Noske,Leipzig,1918))中作出的. 克希纳的论文名为“平面铺砌”(On paving the plane),登在《美国数学月刊》75(1968)卷上,页数为839—844. 我们的五边形清单取自D·沙特斯奈德的文章“用全等的五边形铺砌平面”(Tiling the plane with congruent pentagons),见《数学杂志》51卷(1978),29—44页. M·D·希尔锡洪与D·C·亨特宣称等边五边形的清单已经完备,这篇文章也登在《数学杂志》上,见该刊51卷(1978),312页.

2. k相同态铺砌这一问题可以参阅本文作者所写的另一篇文章“由补钉给出的铺砌”(Patch-determined tilings),见《Mathematical Gazette》61卷(1977),31—38;在名为“砌块与铺砌的规定数”(Prescribed numbers of tiles and tilings)的论文里有着哈包思的例子,该文登在《Mathematical Gazette》61卷(1977),296—299页.

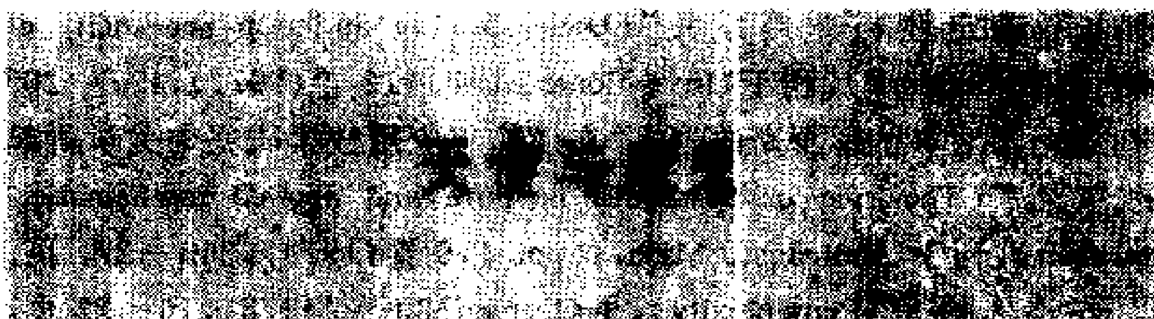
3. 希尔伯特的著名问题(英译文)登载在《Bulletin Of The American Mathematical Society》第8卷(1902),437—479页的“数学问题”一栏上,在题为“希尔伯特问题引起的数学进展”(Mathematical Developments Arising from Hilbert Problems)一文中又作了重印,见《Proc. Sympos. Pure Math.》Vol. 28,(American Math. Soc., Providence,R. I.,1976). 由希许与金兹莱所写的书名叫《平面剩余》(《Flächenschluss》,Springer-Verlag,Berlin-Göttingen-Heidelberg,1963). 有关等面铺砌的近期研究可参看本文作者的文章“81种平面的等面铺砌”(The eighty-one types of isohedral tilings in the plane),登载在《Math. Proc. Cambridge Philos. Soc.》82卷(1977),177—196页上.

4. 扩展定理的一个证法将发表在下文要提到的、本文作者所写的一本书上. 希许问题可参阅他的著作“正则镶嵌问题”(Reguläres Parkettierungsproblem, Westdeutscher Verlag,Köln-Opladen,1968).

5. 有关非周期性铺块的材料，见 R. M. 罗宾孙的文章“平面铺砌的不确定性与非周期性”(Undecidability and non-periodicity of tilings of the plane)，刊于《Inventiones Math.》12 卷(1971)，177—209 页，也可参看 R·彭罗斯的论文“在纯粹与应用数学研究中美学所起的作用”(The role of aesthetics in pure and applied mathematical research)，刊于《Bull. Inst. Math. Appl.》10 卷(1974)，266—271 页，以及“迷人的 5”(Pentaplexity)，见《Eureka》39 卷(1978)，16—22 页，这方面的最新资料即系本文中引用过的马丁·加德纳的文章，在本文作者的一本即将问世的书中将有更详尽的阐述. 非周期性与铺砌问题的联系在上面已引用过的罗宾孙的论文中作了论述. 在王浩的“用模式识别证明定理Ⅱ”(Proving theorems by pattern recognition Ⅱ)中也曾谈及. 该文刊登于《Bell System Techn. Journal》40 卷(1961)，1—42 页.

6. 伏特堡的论文名为“围绕一个平面区域的全等形分解”(Zur Zerlegung der Umgebung eines ebenen Bereiches in kongruente)，见《J.-Ber. Deutsch. Math.-Verein.》46 卷(1936)，229—231 页. 还有另一篇论文“平面区域的螺旋状全等形分解”(Zur Zerlegung der Ebene in kongruente Bereiche in Form einer Spirale)，同上刊 47 卷(1937)，159—160 页. 戈德堡对螺旋形铺砌结构所作之解释见于《Scripta Math.》21 卷(1955)，253—260 页，其论文题目是“中心镶嵌”(Central tessellations). 由本文作者所撰写的一篇较短的文章“螺旋式铺砌与‘多才多艺’的砌块”(Spiral tilings and versatiles)，登在《Mathematics Teaching》88 卷(1979)，50—51 页上.

以上列举的仅仅是平面铺砌问题上已发表的、浩如烟海的文献中的极小一部分. 进一步的信息与问题将在本文作者的一本新著《铺砌与模式》(Tilings and Patterns)中出现，该书不日将由旧金山市的 W. H. Freeman and Company 出版.



● 多伦多大学

□ H・S・M・考克塞特(H. S. M. Coxeter)

大约四十年以前,本人与阿伯拉罕・辛可夫(Abraham Sinkov)写了两篇基本类似的文章,研究由两个生成元S,T及其交换子$S^{-1}T^{-1}ST$的周期所决定的群〔请参看 Coxeter 1936 年一文与 Sinkov 1936 年一文〕,他们绝未梦见,二十年之后,M・C・埃歇尔竟会(无意识地)利用这些群作为一个精雕细刻的球以及四件其他艺术品的对称群〔见 Escher, 1971, 图 112, 115, 226, 235, 244, 247; Mac Gillavry, 1976, p. 18〕. 承荷兰海牙市海牙博物馆埃歇尔基金会的慨然许诺,我们在这里复制了这些艺术品.

欧几里得平面上的模式

在普通的墙纸中,图案的基本主题经由两个方向的平移不断重复着,这与对称群 p1 的性态协调一致. 理论上,更有趣的模式可通过其他对称操作而获得,例如,一个半转(旋转 180°)将使主题上、下颠倒(使字母 b 变成字母 q),一个反射将把左手变作右手(b 变成 d 或 p). 半转是周期为 2 的旋转,但我们也可以利用周期等于 3,4 或 6 的旋转. 考虑这些"等距要素"的一切可能组合,E・S・费道洛夫(E. S. Fedorov)在 1891 年证明了不多不少、恰有十七种平面对称群,其

中也包括两个方向的平移. 在十七个群中，人们无意地发现，已有十一个群被摩尔人在很久以前用于阿尔亨布拉宫的装饰图案. 这十一个群中有几个群，再加上另外的五个群都已被非洲的巴库巴(Bakuba)与贝宁(Benin)部族(均在萨哈拉沙漠以南)在他们的陶器、织物与筐篮中用上了〔参看 Crowe，1971；1975 年的文章〕. 最后的一个群称为 p31m，在中国的一件工艺美术品上也发现了其模式〔请参阅 Fejes Tóth 1964，p. 40 的彩色插页 Plate Ⅱ. 1〕.

埃歇尔的模式显得更为有趣，因为它们的图案主要是一些动物，

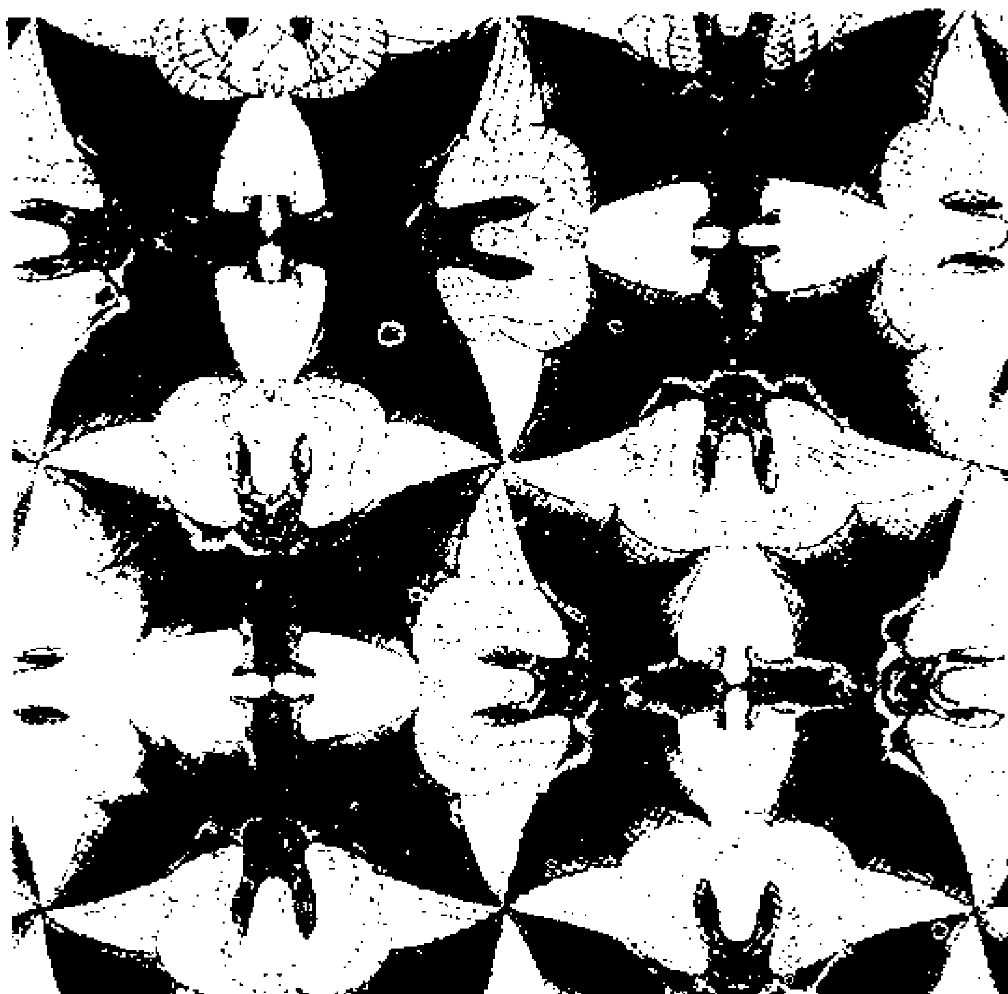


图 1

埃歇尔的素描"天使与魔鬼".

经过精心设计，正好可以铺满平面，不留下一丝空隙．例如，在图1中，平面上的任一点或者属于天使，或者属于魔鬼，或者属于它们的分界线，三者必居其一．模式被视为连续的，因而它可以铺满全平面．围绕图案上的每一个特殊点(四位天使的翼尖与四个魔鬼的翼尖正好在那里相交)，此模式具有"四分之一转"(旋转$\frac{\pi}{2}$)的对称性．对称要素中还包括由某些水平线与垂直线所作之反射，以及所有上述旋转与反射的乘积．更加扼要地说，整个对称群是由一个周期为4的旋转S与一个反射T来生成的．例如，若T是一个垂直镜面的反射，S是一个"四分之一转"，其旋转中心尽可能接近那面反射镜，则S即可把T变换为反射$T_1=S^{-1}TS$，从而得出水平镜面反射的效应．旋转S^k的整数幂($k=0,1,2,3$)●将能把T转换为对正方形各个边所作之反射$S^{-k}TS^k$．

旋转S^k可形成由S生成的四阶循环群C_4．两个反射T与T_1生成四阶二面体群D_2，其中乘积T_1T生成一个子群C_2(因为半转T_1T的周期为2)．

关系式　$S^4=1$与$T_1^2=T^2=(T_1T)^2=1$分别为群C_4与D_2的抽象定义或描述，由生成元S与两个生成元T_1与T所能满足的任何关系式都可以从这些简单关系式推出来．若把T_1改写为$S^{-1}TS$，我们可以推导出以下关系式：

$$S^4=T^2=(S^{-1}TST)^2=1$$

作为由旋转S与反射T所生成的无限群的一个描述．

这一无限群**p4g**〔见Coxeter与Moser1972，第47页〕乃是群$[l^+,2p]$的一个特例$[4^+,4]$．

该群的描述为

$$S^l=T^2=(S^{-1}TST)^p=1 \quad (l\geqslant 2, p\geqslant 1).$$

这里S是一个周期为l的旋转(即旋转角为$2\pi/l$)，而T是一个反

● 译者注：当然要与S^{-k}一起作成共轭变换$S^{-k}TS^k$才行．

射，其反射镜处于这样一种位置，以使得 T_1T（其中 $T_1=S^{-1}TS$）的周期为 p. 换言之，S 的乘幂将把 T 的反射镜转变为一个正 l 边形的各条边，而其相邻的两边（例如 T_1 与 T 的反射镜）形成一个角 π/p. 只要平面可以被正 l 边形（围绕每一顶点有 $2p$ 条边在此相接）铺砌，这样的群 $[l^+,2p]$ 必然会出现. 此类铺砌可记为 $\{l,2p\}$〔请参看 Coxeter 与 Moser，1972，p. 102〕，例如，{4，4}就是常见的正方形铺砌纸（所谓"方格纸"模式）. 欧几里得平面上唯一可能的其他模式是{3，6}，见图 2 的粗黑线.

埃歇尔重新发现费道洛夫对称群中绝大多数成员，这件事给予结晶学家卡罗林·麦克·吉拉夫利（Caroline MacGillavry）〔见 1976，彩图 8〕印象至深，但她同时也指出，埃歇尔也像非洲部族人一样，遗

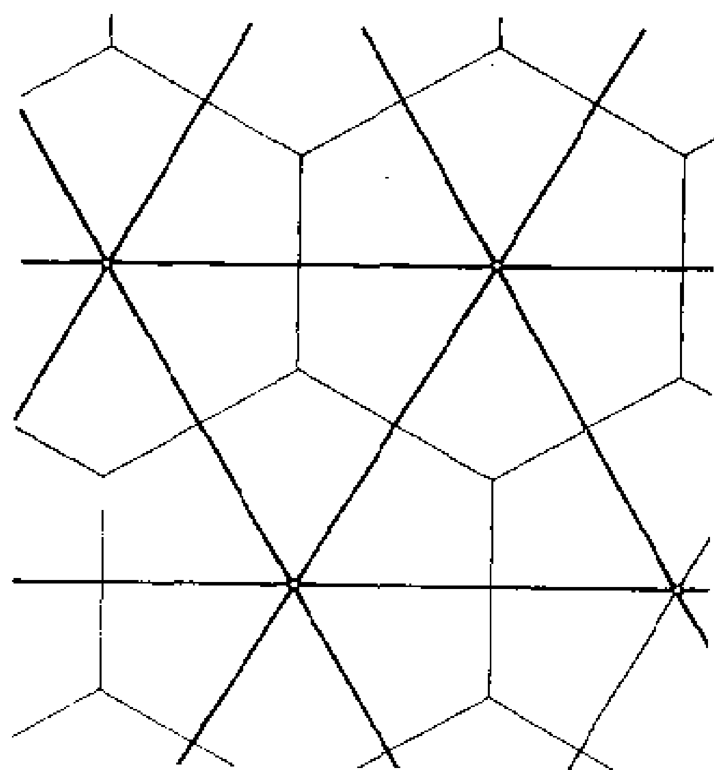


图 2

对偶镶嵌图{3，6}（粗黑线）与{6，3}（细线）的一个片断.

漏了群

$$p31m \cong [3^+,6]$$

(在〔Coxeter,1969,p. 413〕以及〔Coxeter 与 Moser,1972,pp. 49,136〕两篇论文里,它被错误地命名为 p3m1). 应她的请求,埃歇尔重新创作了一幅红色蜜蜂与黄、绿色的黄蜂图(见彩色插图Ⅳ)来弥补这个缺陷. 图中,背景镶嵌图{3,6}的各边显然是这些昆虫的对称直线. 这些边被其对偶镶嵌图{6,3}的各边所垂直平分(见图 2 中的细线),而{6,3}镶嵌图的各个顶点则是三只蜜蜂的"肘部"与三只黄蜂的"肘部"的交会点. $[3^+,6]$群的生成元 S 是围绕这类点、转角为 $2\pi/3$ 的旋转,我们很容易验证下列关系式为真:



图 3

埃歇尔的"鱼球图".

$$S^2 = T^p = (T_1T)^2 = 1,$$

其中 $T_1 = S^{-1}TS$.

我们也可以把记号$[2^+,2p]$的意义加以推广，使之包括

$$[2^+,\infty] \text{与} [\infty^+,2],$$

这就把七种带饰[1]群中的第五与第六个群〔见 Coxeter，1969，p. 48〕也收编了进去. 前者由一个“半转”与一个“反射”所生成，它是正弦曲线与下列带饰

…… V Λ V Λ V ……

的对称群；后者则由一个平移与一个反射（按反射镜面的方向进行平移）所生成，是下列带饰

……D D D D D……

的对称群.

雕 琢 之 球

如果我们用非欧平面来取代通常的平面——用球面或双曲面均可，则欧氏镶嵌图的一种更为有趣的推广即随之产生.

球面可以看作是平面，其上的直线乃是大圆. 这一看法来自阿普瓦法（Abû’l Wafâ）（940—998 年）〔参阅 Woepke，1855，pp. 352—357〕. 围绕球面直径的转动可以看作是围绕此直径与球面的两个交点的任一个所作之转动. 内接于球的正四面体{3，3}有着绕其顶点、作周期为 3 的旋转对称性，这些旋转生成了阶数为 12 的四面体群，它可以用记号 A_4 来表示，因为它是四次交代群，即 4 个顶点的偶排列所成之群. 在同一个球内还可内接四个正四面体，因此全部 20 个顶点属于一个正十二面体{5，3}〔参看 Coxeter 与 Moser，1972，p.

[1] 译者注：原为建筑学名词，其意思是指柱的中楣，今根据我国著名数学家段学复先生的看法，译为“带饰”. 详见其著作《对称》一书.

35]. 对 A_4 群追加一个周期为 5 的旋转以后，我们即可得出阶数为 60 的二十面体群，它可记为 A_5，因为它是五次交代群，即五个四面体的偶排列所成之群.

埃歇尔在其制作的工艺品，一个雕刻球面上用 A_4 作为其对称群（见图3），而在另一个球面上采用了A_5（见图4）. 在后一情况，十



图 4

花团锦簇的多面体.

二朵花呈螺旋状，有点像海螺或长春花，花瓣的顶尖便是十二面体的顶点. 事实上，这只球的设计原理完全是依据上面提到的复合多面体(见图 4a)，它可记为

$$\{5,3\}[5\{3,3\}]\{3,5\},$$

因为它的二十个顶点属于一个正十二面体{5,3}，而其二十个面则与对偶十二面体{3,5}的各个面相重合.

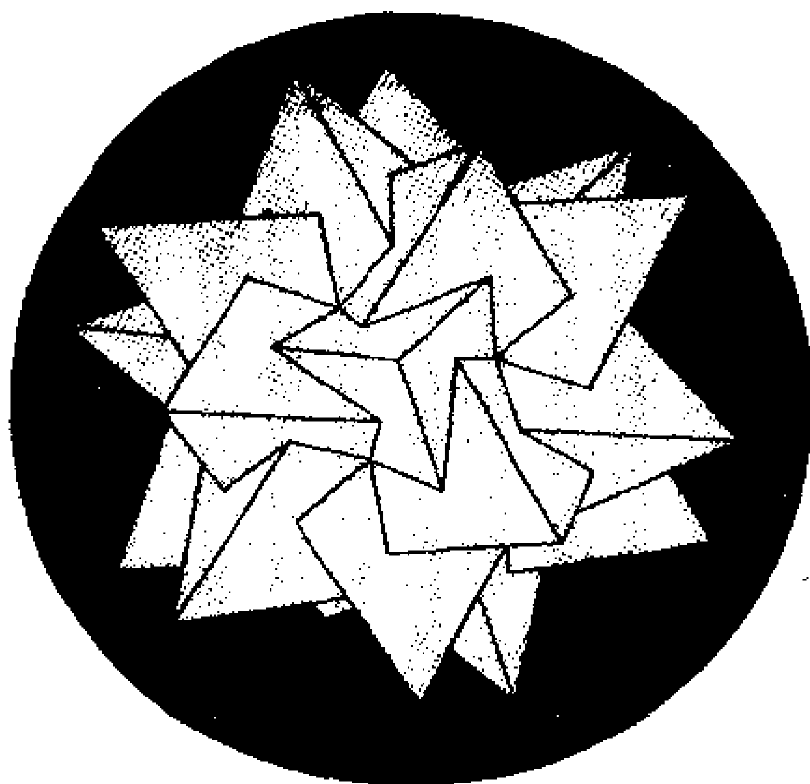


图 4a

J·F·配德利(J. F. Petrie)所描绘的五个四面体.

与我们所讨论的群$[l^+,2p]$更具有直接联系的是他的其他雕刻球作品(见图 5)，因为它具有周期为 3 的旋转对称性，而这个旋转 S 的中心是三位天使的翼尖与三个魔鬼的翼尖的交会点. 它也具有反射对称性，该反射 T 的反射镜面是三个相互垂直的平面之一. 于是 S 与 T 生成了群$[3^+,4]$，而作为其背景的球面镶嵌图{3,4}则是

一个八面体，其各个面是球面被这三个对称平面所截出的球面三角形. $[3^+, 4]$有时称为五角十二面体群，因为它也是一种黄铁矿晶体的对称群，而黄铁矿晶体则是不正则的十二面体. 若用抽象群的说法，这个阶数为 24 的群乃是 $A_4 \times C_2$ 的直积〔参阅 Coxeter 与 Moser, 1972，第 3 与 39 页〕.

事实上，作为其定义的关系式

$$S^3 = T^2 = (S^{-1}TST)^2 = 1$$



图 5

精雕工艺球"天使与魔鬼".

可为下列置换

$$S = (abc), T = (ab)(cd)(ef)$$

所满足，它们将作成 $\mathbf{A}_4$ 与 $\mathbf{C}_2$ 的直积. 其中，$\mathbf{A}_4$ 由

$$(abc) = S,\ (ab)(cd) = STSTS$$

生成，而二阶群则由

$$(ef) = (ST)^3$$

生成.

为了完整起见，我们也应当提到平凡肤浅的两个群

$$[2^+, 2p] \cong \mathbf{D}_{2p} \text{与} [l^+, 2] \cong \mathbf{C}_l \times \mathbf{C}_2$$

（阶数分别为 $4p$ 与 $2l$），它们很像群$[2^+, \infty]$与$[\infty^+, 2]$，仅不过相应的带饰是被覆在一个圆柱上的. 换句话说$[2^+, 2p]$是一个 p 多边形反棱柱的对称群〔见 Coxeter, 1969, p. 149〕.

圆 周 极 限

1958 年，我送给埃歇尔一件复制品，其中有幅插图极像图 6. 他复信表示感谢并说了下面一番话："（插图使我）深感惊讶. 长期以来，我一直对那样的主题抱有强烈兴趣，它们越变越小，以迄于无限小的极限. 如果极限是模式中央的一个点，则这个问题是较为简单的. 直线形状的极限对我来说也不稀罕，可是我从来没有能够作出一个模式，像你的图形所表现的情况，每个"墨团团"从中心向四周越变越小，直至最外面的圆周极限. 我试图找到图形的几何作法，但我只能求出最大内圆的圆心与半径. 如果您能简单地解释一下其他各圆的作法（其圆心由外边逐渐接近以至于极限），我将喜悦万分并感激不尽！除了这一个图形之外，还有没有可以达到圆周极限的其他系统？"

在复信中，我告诉埃歇尔：$\{4,6\}$与$\{6,4\}$仅仅是无限多种正则镶嵌图$\{p,q\}$中的两种而已，图中的每个顶点都有 q 个配合好了的、完

全合同的正 p 边形. 如认为 p 与 q 的数值过大，在球面或欧氏平面上不易描出相应的镶嵌图，那就需要利用双曲面. 于是，正p边形就会

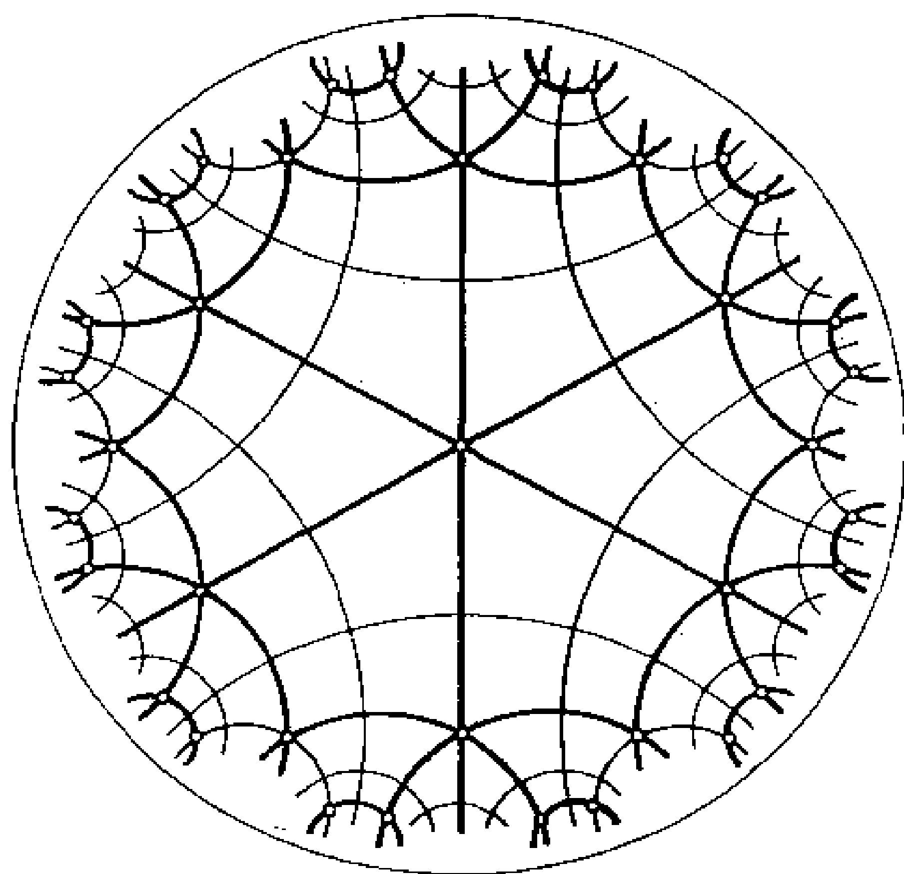


图 6

对偶镶嵌图{4,6}(粗黑线)与{6,4}(细线)的一个片断.

具有一个较小的顶点角. 在庞加莱(Poincaré)的一种模型中，双曲面上的"直线"表现为圆弧，它与欧氏平面上所作之境界图 Ω 正交. 天使们都得到了如实的描绘，然而距离却遭到扭曲，而境界圆本身的点则在无穷远处. 虽然图 6 中粗黑线所形成的正则四边形"由中央向四周越变越小，以迄于外面的圆周极限"，我们还是要佯装不知，把这些四边形都看做是正则与合同的，我们正是通过此种想象力的扩展，而进入了双曲几何的精神世界.

埃歇尔的素描本表明，在他完成其杰作“圆周极限Ⅳ”（见图 7）之前，他曾勤奋地钻研过这些概念．“圆周极限Ⅳ”的对称群（在双曲几何意义下）为$[4^+,6]$：

$$S^4 = T^2 = (S^{-1}TST)^3 = 1.$$



图 7

埃歇尔的圆周极限Ⅳ．

其中的生成元 S 是（类似于图 1 中的）绕着一个点所作的四分之一转，而此点则是四位天使的翼尖与四个魔鬼的翼尖的交会诸点中的一个，换言之，即为{6，4}镶嵌图（图 6 中通过细线表示）中的一个顶

点. 另一生成元 T 则是对图 6 中一条粗黑线(图 6 包含{4,6}图中无限多条边)的反射. 如果 T 的反射镜面是通过中心的直线之一,则 S 将把它变换为一个粗黑线的弧,而双曲反射 $T_1=S^{-1}TS$ 将在庞加莱模型中表现为对包含那条弧的圆所作之反演. 图 7 有异于图 1 的是,旋转 T_1T(旋转中心是天使的脚尖)的周期为 3,而不是 2.

埃歇尔利用了类似的群[3^+,8]来创作他的"圆周极限 Ⅱ"(见图 8),它在数学上同样饶有趣味,虽然布鲁诺·恩斯特(Bruno Ernst)〔见1976,p. 109〕开了一个玩笑,把它打发掉了. 所有十字形的中心

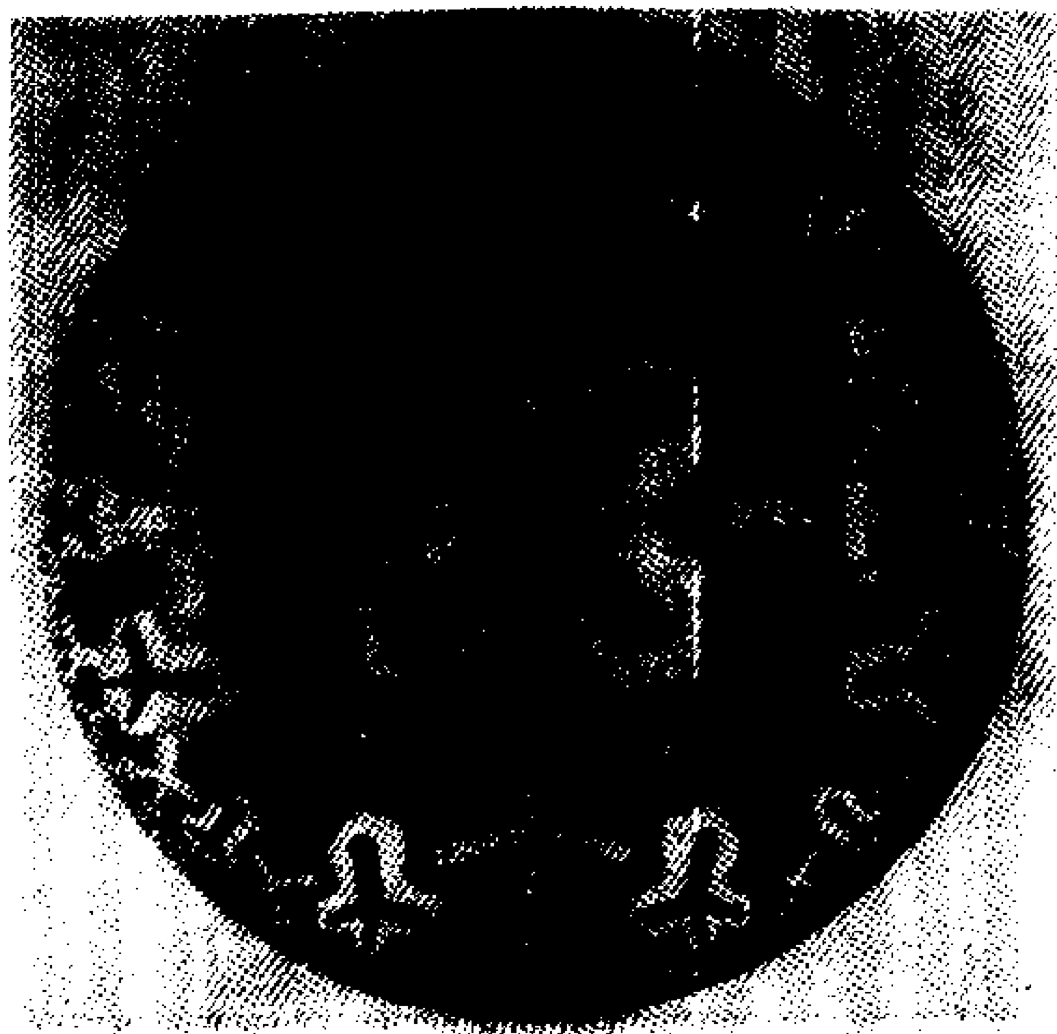


图 8

埃歇尔的圆周极限 Ⅱ.

都是背景镶嵌图{3,8}(见图 9)中的顶点.三种颜色交会处的点就是{3,8}构形中三角形面的中心,而在其对偶构形{8,3}中则为顶点〔见 Coxeter,1979,p. 23 的图 5〕.另一方面,一种颜色十字架的中心是四边形镶嵌图{4,8}的顶点(在图9中,与黑、白、灰色十字形相一致,

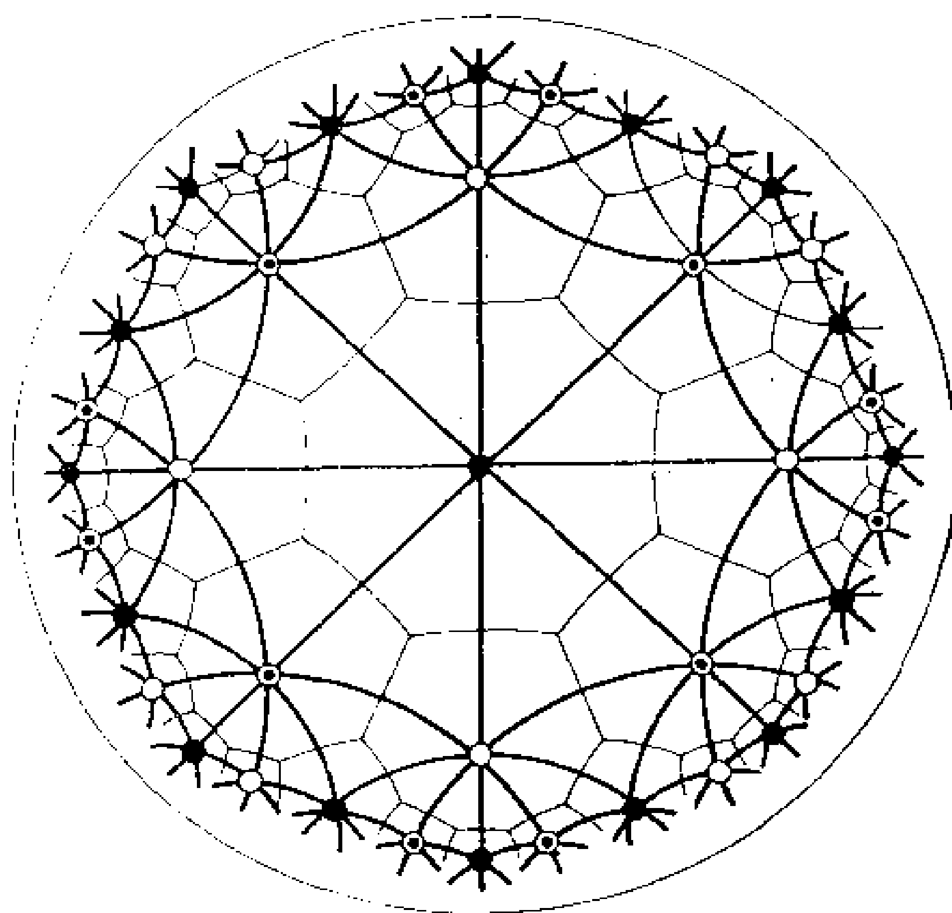


图 9

对偶镶嵌图{3,8}(粗黑线)与{8,3}(细线)的一个片断.

与之对应的构形{3,8}的顶点也分别标以黑点、白点与小圆点).就此种意义来说,可以说埃歇尔事先已经预料到了我的发现.那便是正则复合镶嵌图

$$\{3,8\}[3\{4,8\}]2\{8,3\}$$

〔参看 Coxeter 1964 的论文,pp. 156—157〕,其中包括 3 个叠置的镶

嵌图{4,8},其顶点属于单个构形{3,8},而其面的中心数则与对偶构形的面数(每一个都用上两次)一样多.

结　论

对任意两个整数 l 与 $p(l>2,p>1)$ 存在着一个群 $[l^+,2p]$,它由周期为 l 的旋转 S 与周期为 2 的反射 T 所生成,而交换子 $S^{-1}TST$ 则是周期为 p 的旋转.有关的"面"分别为球面、欧氏平面或双曲面,取决于数 $(l-2)(p-1)$ 是小于 2、等于 2 还是大于 2. 有限群 $[3^+,4]$(阶数为 24)是埃歇尔的作品"球面上的天使与魔鬼"中的对称群.他还利用了两个欧氏群 $[4^+,4]$ 与 $[3^+,6]$.有趣的是,在满足

$$(l-2)(p-1)>2$$

的无限多种双曲群 $[l^+,2p]$ 中,他本能地选中了两个最简单的群:$[4^+,6]$ 与 $[3^+,8]$.

参考文献

1 Coxeter, H. S. M. 1936. The groups determined by the relations $S^l=T^m=(S^{-1}T^{-1}ST)^p=1$ *Duke Math. Journal.* 2: 61—73.

2 ________. 1964. Regular compound tessellations of the hyperbolic plane. *Proc. Royal Soc. London* A 278: 147—167.

3 ________. 1969. *Introduction to Geometry.* 2nd ed. New York: Wiley.

4 ________. 1979. *The non-Euclidean symmetry of Escher's picture 'Circle Limit III'*. Leonardo 12: 19—25, 32.

5 ________ and Moser, W. O. J. 1972. *Generators and Relations for Discrete Groups* 3rd ed. Berlin: Springer.

6 Growe, D. W. 1971. The geometry of African art I. *J. Geom.* 1: 169—182.

7 ________. 1975. The geometry of African art II. *Historia Math.* 2: 253—271.

8 Ernst, Bruno. 1976. *The Magic Mirror of M. C. Escher*. New York: Random House.

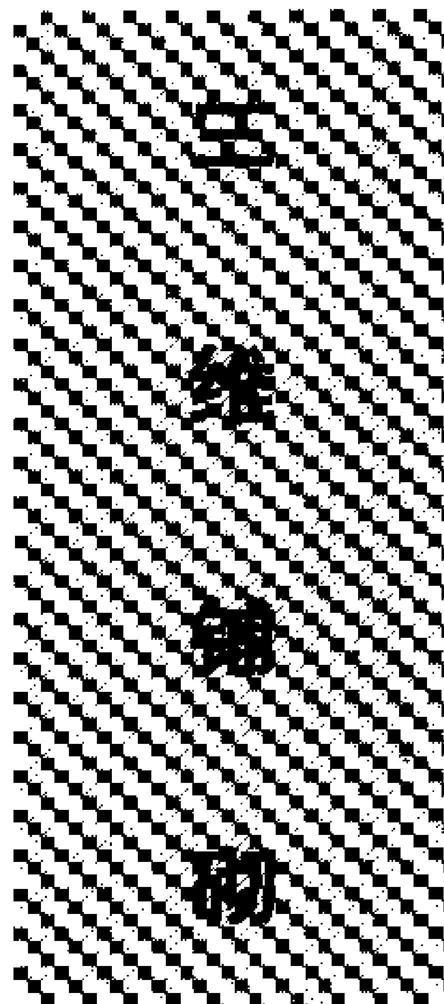
9 Escher, M. C. 1971. *The World of M. C. Escher*. New York: Abrams.

10 Fejes Tóth, L. 1964. *Regular Figures*. New York: Pergamon.

11 MacGillavry, Catoline. 1976. *Fantasy and Symmetry—The Periodic Drawings of M. C. Escher*. New York: Abrams.

12 Sinkov, Abraham. 1936. The Groups Determined by the Relations $S^l = T^m = (S^{-1}T^{-1}ST)^n = 1$. *Duke Math. Journal* 2: 74—83.

13 Woepke. F. 1855. Recherches sur l'histoire des sciences mathématiques chez les orientaux, d'après des traités inédits arabes et persans. *J · Asiatique* 5: 309—359.





● 奥伯恩大学

□ D·G·霍夫曼(D. G. Hoffman)

要在一个边长15英里的正方形行政区中，划出四块长与宽各为7英里与8英里的农地，应该怎样做？（在对下页图1(a)的答案偷偷地瞥上一眼之前，最好请你自己先尝试一下.）

请注意行政区的面积为 $15^2=225$ 平方英里，而每块农地的面积是56平方英里，因而还余下 $225-4\times56=1$ 平方英里.

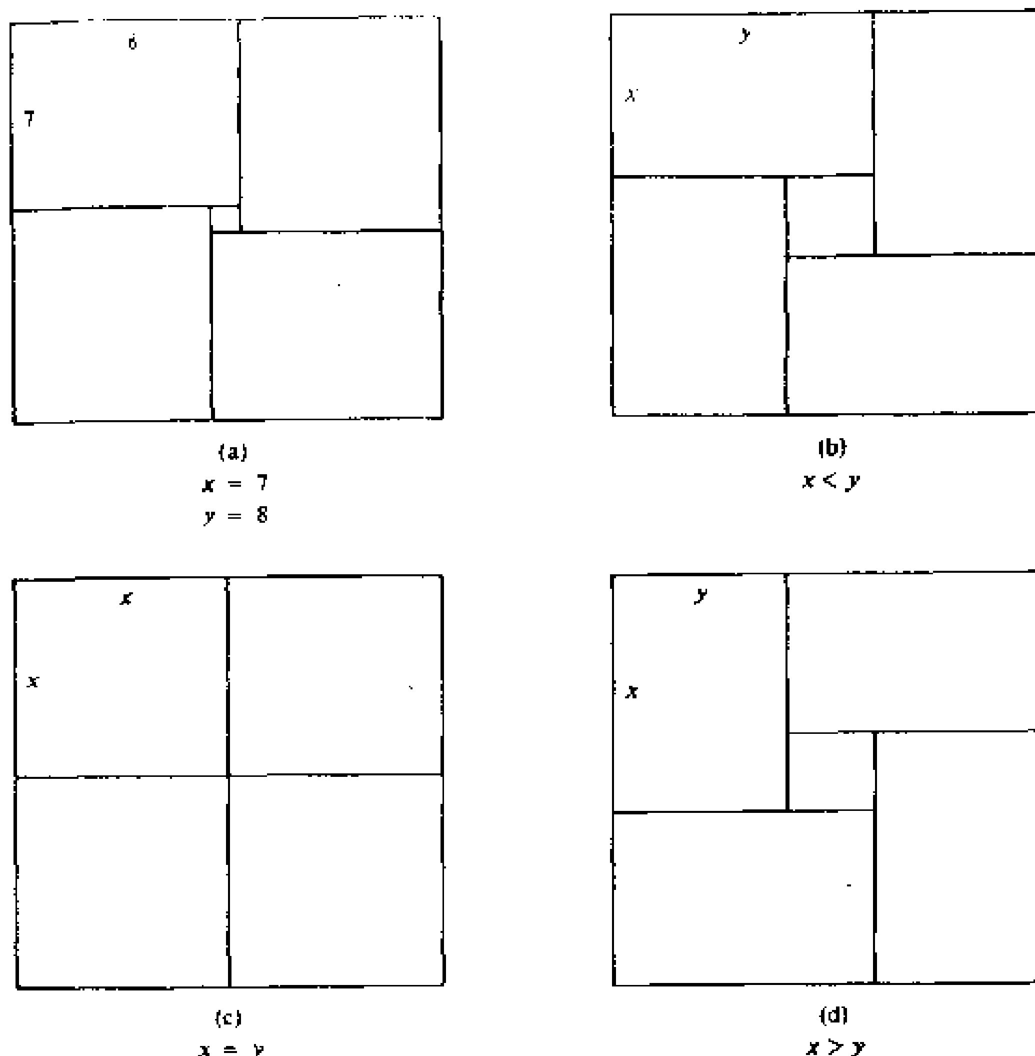
现在有一个更一般的问题. 设 x、y 为正数，则四块矩形农地（每块大小都是 x 英里 $\times y$ 英里）能否适当置入边长为 $x+y$ 英里的一个正方形行政区？显然这一行政区的面积是 $(x+y)^2$ 平方英里，而每块农地的面积是 xy 平方英里. 所以，除非

$$1. \qquad 4xy \leqslant (x+y)^2$$

我们是没有可能解决本问题的. 换句话说，如果问题有解，则四块农地的总面积不能超过行政区的面积，即不等式1必须得到满足.

这个更一般的问题确实有解，我能肯定，如果你能解决第一个问题，即 $x=7$，$y=8$ 的情况，那么你一定也能解决它. 下面的图1(a)，给出了第一个问题的解. 而(b)、(c)、(d)则分别根据 x 与 y 的相对大小，给出了一般问题的解.

两数 x、y 的算术平均数是其和数的一半，即



(a)
$x=7$
$y=8$

(b)
$x<y$

(c)
$x=y$

(d)
$x>y$

图 1

2. $$\text{A. M.}=\frac{1}{2}(x+y).$$

当 x、y 都是正数时，还有 x、y 的另一种平均数，它有时是很有用处的．它叫做 x、y 的几何平均数，是指其乘积的平方根：

3. $$\text{G. M.}=\sqrt{xy}.$$

例如，若 $x=y$，则 A. M. $=\frac{1}{2}(x+x)=x$，而 G. M. $=\sqrt{xx}=x$，因而当 $x=y$ 时，这两个平均数相等．然而，当 $x=4$，$y=16$ 时，A. M. $=10$，而 G. M. $=8$．一般来说，几何平均数永不大于算术平均数，即

4. $$\sqrt{xy}\leqslant\frac{1}{2}(x-y).$$

利用上述“四块农地”问题，可以得到一种简易证法. 图 1 表明该问题确实有解，因此不等式 1 必然成立. 我们在 1 式的两端各取平方根，得出 $2\sqrt{xy}\leqslant x+y$，用 2 除此式的两端，便得到不等式 4，结束了我们的证明.

把一些物体妥帖地装入一个容器的趣题称为装箱问题. 这类趣题对马丁・加德纳的读者、研究数学游戏的学者以及走私贩子们都是很熟悉的.

上面这个例子给我们的教益是：一个装箱问题的任何解法提供了一个不等式的证法. 倘若所有的物体都能妥帖地装入容器，则它们的总面积（或体积）必定不大于容器的面积（或体积）. 可能反过来说也是对的. 也就是说，如果我们从一个已成立的不等式出发，能否由此“造出”一个有趣的装箱问题. 下文给出一个实例.

x、y、z 三数的算术平均数是它们和数的三分之一，即

5. $$\text{A.M.}=\frac{1}{3}(x+y+z).$$

三个正数 x、y、z 的几何平均数是它们乘积的立方根：

6. $$\text{G.M.}=\sqrt[3]{xyz}.$$

有关两个数的不等式 4，也可推广到三个数的场合，即：

7. $$\sqrt[3]{xyz}\leqslant\frac{1}{3}(x+y+z).$$

为了从此式导出一个装箱问题，我们将不等式的两端乘上 3，并各自立方，其结果是：

8. $$27xyz\leqslant(x+y+z)^3.$$

我们注意到 xyz 是一块尺寸为 $x\times y\times z$ 的砖块的体积，而 $(x+y+z)^3$ 则是边长为 $x+y-z$ 的立方体之体积. 于是得出了如下的包装问题：

> 每块砖的大小为 $x\times y\times z$，请问，27 块一样大小的这种砖，能否妥帖地装入一只边长为 $x+y+z$ 的立方体箱子？

当然，如果此问题有解，则该解法就能证明 8 式（从而也可推出 7 式）是一个的确成立的不等式.

上面我们已经提到，对任意正数 x、y、z，7 式都能成立. 然而这一事实本身并不能保证砖块可以装入箱子，而仅不过说明它们的体积之和不大于箱子体积而已.（我的钓鱼竿之体积小于我的公文包之体积，可是钓鱼竿却装不进我的公文包！）

说到这里，你也许会真的去弄一组砖块（27 块）来动手解决这个趣题. 我可要劝告你怎样去选择 x、y、z 的适当尺寸. 选择 x、y、z 时有一个基本问题. 有些 x、y、z 所提供的趣题简直是易如反掌的！例如，$x=1$，$y=1$，$z=100$ 引人发噱. 27 根手杖（每根尺寸为 $1''\times 1''\times 100''$），当然可以装进 $102''\times 102''\times 102''$的大盒子，留下的空隙足够放得进一只低音大喇叭. 另一极端情形是 $x=y=z$，这时 27 块砖全是立方体形状，可以服服帖帖地装进盒子. 为了保证避免这种肤浅的情况，必须确保长、宽、高三个尺寸都不相同，而且两个较大数字之和要小于最小数的三倍.

也就是：

9. $$0<x<y<z,$$

10. $$y+z<3x.$$

如果你搞到的一组砖块不满足 9 式或 10 式，那就有可能（而且是极有可能）轻易地找出问题的解.

除了 9 式与 10 式，选择 x、y、z 的合适尺寸所需之唯一其他因素是要看你手头有什么材料可以利用，制造起来是否方便. 不存在满足 9 式与 10 式的一组 x、y、z，由它们所产生的包装趣题将会比任何一组别的数值更为容易或更为困难，可能只有一个例外. 这就是说，$y=\frac{1}{2}(x+z)$所产生的趣题也许要比较难一些！这里有一个很好的例子：$x=4$，$y=5$，$z=6$. 我们要问：27 块 $4\times 5\times 6$ 的砖头能不能妥帖地装进一只 $15\times 15\times 15$ 的箱子？

让我向你保证，这个趣题确实有解. 事实上，我听 J·H·康威（J. H. Conway）与威廉·寇特勒（William Cutler）说，本问题恰好有 21 个解，不计算箱子的旋转与反射，当然，这是在满足 9 式与 10 式的前

提下说的.

要找出问题的一个解看来是很困难的,我想只利用铅笔与纸头来解决此问题,试算了好几个钟点.最后,我宣告放弃,并请来了我的朋友戴维·克拉纳,他手中拿了把锯子.他告诉我,当他把锯下来的木块一块块地用沙擦光,并按照下面附图中给出的解法堆放起来时,他是胸有成竹的,连一块木头都未改动过.于是他赢得了荣誉,成为该问题的第一个解决者(请参看次页的一些照片).我已看到人们解

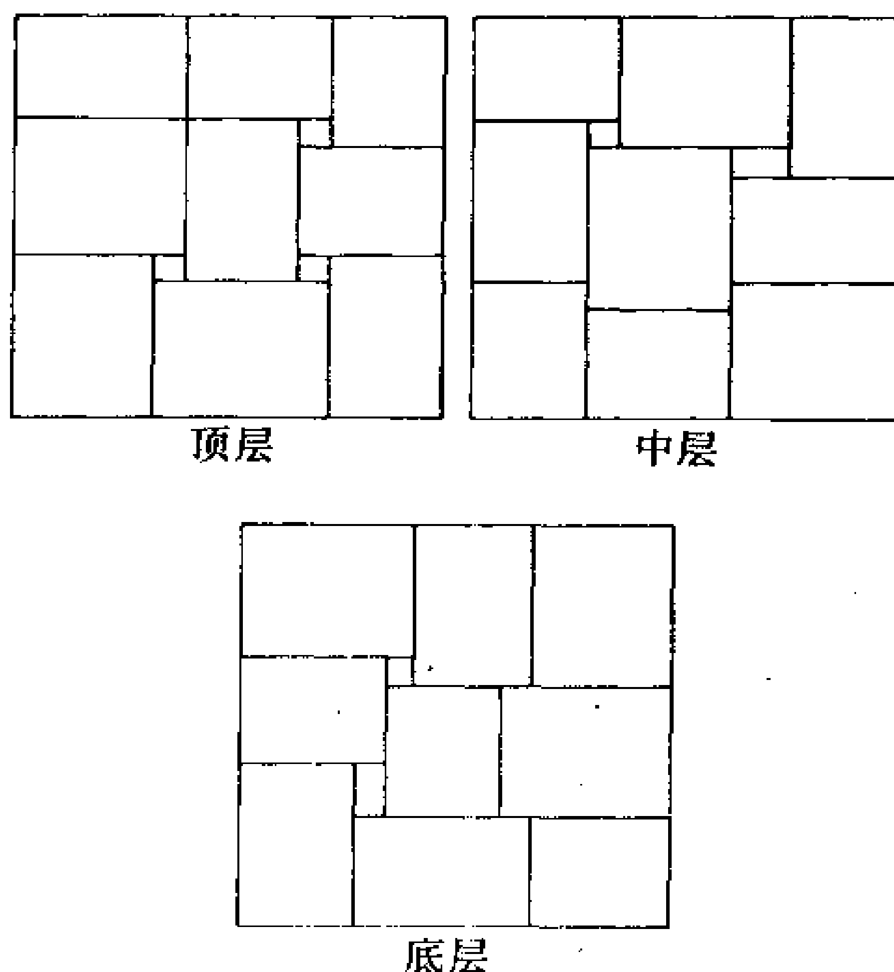


图 2



1



2



3



4

首先造出一套木块，并发现一种巧妙的装箱法的是D・A・克拉纳．上面附图中表明，卡尔・克拉纳(Carl Klarner)怎样把这些木块(每块的大小为7×8×10)一块一块地堆放起来．拍摄这些照片的人是卡拉・林恩・克拉纳(Kara Lynn Klarner)．

决该问题的时间，短的只有 20 分钟，长的则多达数天. 如果你的手中已经有一套木块，并希望把它们放回到箱子里去（如果你已经做完了游戏，你当然应该这样做），我在图 2 中给出 $x=4, y=5, z=6$ 时的问题解法. 当然，木块的实际大小是无关紧要的，你可以把图 2 利用于任意满足 9 式与 10 式的 x、y、z，正如图 1(b)可适用于任意的 $x<y$.

27 块木块排列为三层，每层 9 块，图 2 是每一层的截面图. 让我们仔细考察这个趣题. 我们已假定 x、y、z 是满足 9 式与 10 式的一些数.

边长为 $x+y+z$ 的一个完整立方体可以经九次锯切而分成 64 块小立方体，其边长为 $\frac{1}{4}(x+y+z)$（见图 3(a)）. 这九次锯割决定了九个平面，我们将把它们称为特殊平面（见图 3(b)）. 九个平面归属于三个集合，其中每个集合有三个相互平行的平面. 在每个集合里，中间的平面与另外两个平面相距 $\frac{1}{4}(x+y+z)$，而后两者又与箱子的外缘相距 $\frac{1}{4}(x+y+z)$，不平行的两特殊平面相交于一直线，这样的直线一共有 27 条，我们把它们称为特殊直线（见图 3(c)），两两都不平行的三个特殊平面相交于一点，这样的点共有 27 个，我们称之为特殊点（见图 3(d)）. 现在设想要用 27 块 $x\times y\times z$ 木块装入这个大立方体.

现在考虑三个集合（每个集合都有三个相互平行的平面）中的一个. 三个平行平面把立方体分为四片，每片的厚度为 $\frac{1}{4}(x+y+z)$. 在我们的装箱中，27 块中的任何一块是否都得以完全躺在这四片中某一片的内部，与三个锯痕中的任意一个都“奇迹般”地不相交呢?

由于 $y+z<3x$，我们有 $x+y+z<4x$，

即
$$\frac{1}{4}(x+y+z)<x.$$

由是得知，我们手中那套木块的最小尺寸 x 要大于每块平板的宽度 $\frac{1}{4}(x+y+z)$. 因此，木块可不像魔术家的漂亮女助手，它们是没

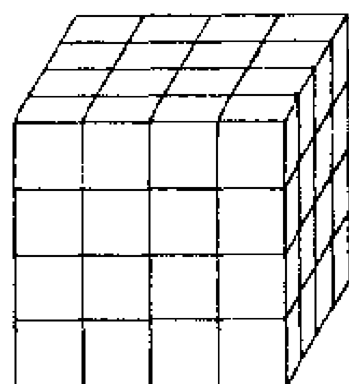


图 3(a)
立方体的锯割.

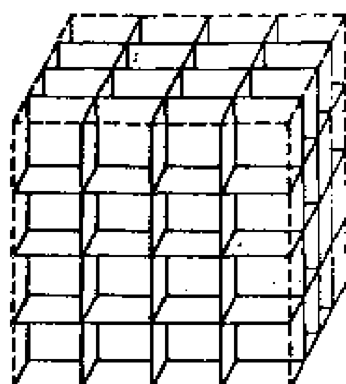


图 3(b)
九个特殊平面.

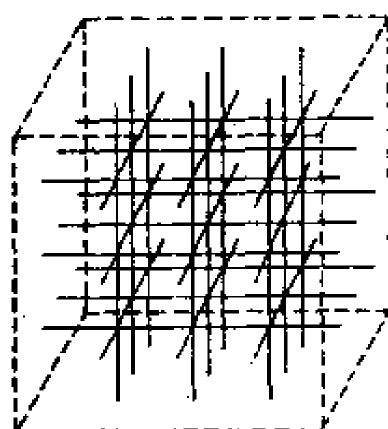


图 3(c)
27 条特殊直线.

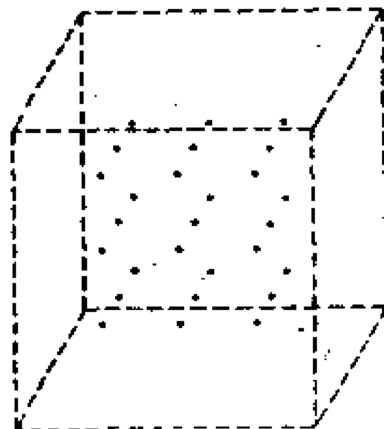


图 3(d)
27 个特殊点.

法避开三个平行锯割的.

现在把我们业已证明的事实简要地归纳一下.在装箱过程中,每一块木块在每个方向上都至少要被三个特殊平面中的一个所切割,在长、宽、高三个方向都是如此.其后果是,这样三个特殊平面的交点必然位于木块的内部.我们已经证明,在装箱过程中,27 块木块中的任意一块必定至少有一个特殊点位于其内部,但是一共也只有 27 个特殊点.因此,每一木块必然是正正好好有一个特殊点在其内部;每

一特殊点必然正好是在某一个木块的内部(顺便说一句,我们也已证明了,在箱子中不可能装入 28 块或更多的木块,因为与之周旋的特殊点不够数).

每一条特殊直线上有着三个特殊点,它们都"躺"在包含这三个特殊点的木块内部.那么,是否有可能像图 4 那样,存在着空隙呢?

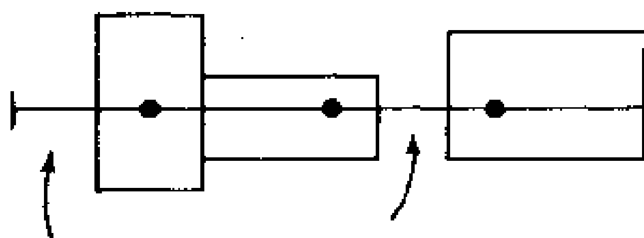


图 4

特殊直线上的"空隙".

我们能够证明不可能有此种空隙存在.让我们来算算看,在装箱过程中,对每条特殊直线来说,其全长 $x+y+z$ 中,究竟有多少是落在木块内部的.让我们把这 27 个数(27 条特殊直线中,每条直线都各有一数)统统加起来,并以记号 l 表示其和数.27 个木块都各有一个特殊点位于其内部,而这个特殊点有三条特殊直线从三个正交方向通过它.因此,每一木块至少必须"吞吃"掉特殊直线的 $x+y+z$ 个单位长.于是有 $27(x+y+z)\leqslant l$,另一方面,27 条特殊直线中每一条的长度是 $x+y+z$,于是 $l\leqslant 27(x+y+z)$,把这两个不等式联系起来看,只可能是 $l=27(x+y+z)$,每条特殊直线必须完全处于具有这三个特殊点的三个小木块内部.特别地讲,像图 4 那样的空隙不可能存在.因此,每条特殊直线看上去都像图5,而$a+b+c=x+y+z$,而且

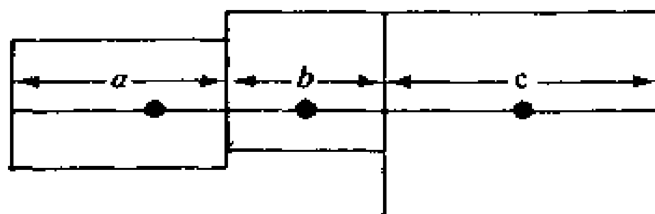


图 5

a、b、c 都是 x、y、z 中的一个，由于 $x<y<z$，因而推知只可能有两种选择，要末是 a、b、c 是 x、y、z 的某一排列，要末是 $a=b=c=y$，另外，除非 $y+y+y=x+y+z$，即 $y=\frac{1}{2}(x+y)$，后一种情况不可能产生.

我们能够证明第二种情形 $a=b=c=y$ 根本不可能发生. 27 条特殊直线中的每一条都被包含此直线的三个小木块分作三段. 因此总共有 27 • 3=81 段. 另外，这 81 段必定正好含有 27 个 x，27 个 y 与 27 个 z，这是因为 27 个木块的每一块都割出这三段，每种尺寸都轮到一段. 上文已指出，任一特殊直线都有一段长度是 y 或者三段全是 y. 但是总共也只有 27 段 y 长，由是得知每条特殊直线必定正好有一段具有 y 长度. 所以 $a=b=c=y$ 这种情形不可能产生.

我们已经证明 27 条特殊直线中的每一条直线都完全处于装箱木块的内部，这些木块把直线分成具有不同长度 x、y、z 的三段，这是一桩很有用场的事实. 譬如说，如果你想用图 6 所示的办法，在箱子底层摆放这九块小木块，你就不可能解决这个装箱难题，你能看得出这是什么道理吗?

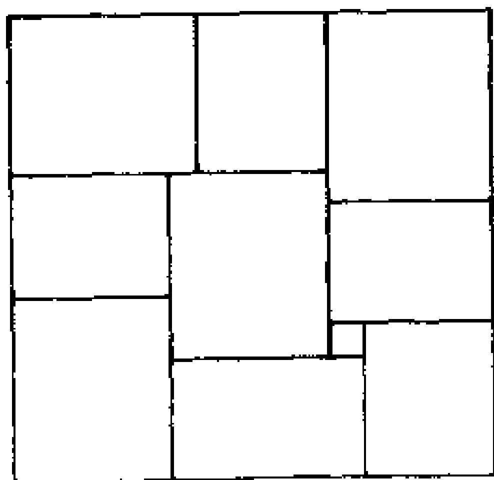


图 6

另一桩有用事实是，没有一个特殊平面的角上可作图 7 那种安排，这是因为不论哪个木块要放到此平面中去时都必须同时与 A 和

B紧紧毗连，然而C却从中作梗.知道上述这些事实肯定有助于这个趣题的最终解决，但仍会遇到许多困难以及一些错误的开始步骤.

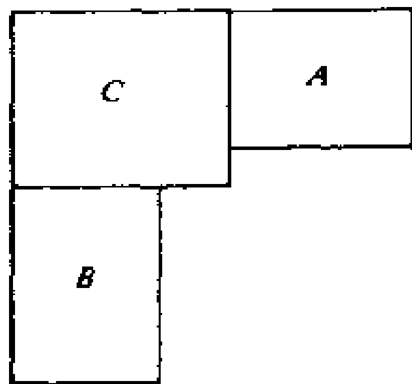


图 7

请注意，九个特殊平面的每一个都含有九个特殊点，因而它穿过九个木块，于是，这一特殊平面将被九个矩形所充填，其中的每个矩形之长、宽为$x\times y$，$x\times z$或$y\times z$，于是，易于证明下列事实：这九个矩形正好是：三个$x\times y$的，三个$x\times z$的.三个$y\times z$的.这方面的三个实例可以参看上面的图2，顺便说一句，我们也已证明了

11. $$3(xy+xz+yz)\leqslant(x+y+z)^2,$$

这是因为，左边是九个矩形的总面积，右边是它们所装填的正方形面积.

如果满足9式与10式，则不计算旋转与反射，一个边长为$x+y+z$的正方形正好有78种方法被三个$x\times y$矩形，三个$x\times z$矩形与三个$y\times z$矩形所充填.

不等式4与7是所谓“算术平均数与几何平均数不等式”的特例.后者断言，若$x_1,x_2,\cdots,x_n$均为正数，则有

12. $$\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}\leqslant\frac{1}{n}(x_1+x_2+\cdots+x_n).$$

左边称为这些数的几何平均数，而右边则是人们熟悉的算术平均数.

为把12式变成一个趣题，先将两边各乘上n，然后各自取n次

幂，其结果是

13. $$n^n(x_1\cdot x_2\cdot\cdots\cdot x_n)\leqslant(x_1+x_2+\cdots+x_n)^n,$$

这里，$x_1\cdot x_2\cdot\cdots\cdot x_n$ 是具有尺寸 $x_1,x_2,\cdots,x_n$ 的 n 维木块的体积，而 $(x_1+x_2+\cdots+x_n)^n$ 是一个 n 维完全立方体的体积，此立方体的每边之长为 $x_1+x_2+\cdots+x_n$，这样的 n^n 个木块能够装入 n 维立方体吗？

让我向那些不习惯于思索四维或更多维的读者们保证，这样一种思考，如果适当地操练一下的话，是决不会像谣传所说的那样，引起相当厉害的大脑损伤的．如果上面的这段叙述使得你害怕了，那么下面将给出一些解释．

正如一个矩形有两维（长与宽），一只皮鞋盒子有三维那样，一个 n 维的木块或箱子有着 n 维，$x_1,x_2,\cdots,x_n$，其中每维的测度 x_i 为一正数．让我们用向量 $(x_1,x_2,\cdots,x_n)$ 来表示一个木块或箱子．例如(3,5)表示一个 3×5 矩形，(4,6,2,6)表示一个 4×6×2×6 的四维“箱子”，而(2,3,9)与(2,9,3)则都表示一个 2×3×9 的普通三维箱子．（在一维，即 $n=1$ 时，(x_1) 表示长为 x_1 个单位的一个线段）．

箱子 $(x_1,x_2,\cdots,x_n)$ 所拥有的“n 维材料”的测度便是数 $x_1\cdot x_2\cdot\cdots\cdot x_n$，即其各个维的乘积．它称为箱子的超体积（广义体积）．这样，对 $n=1$ 来说，广义体积即是长度，对 $n=2$，广义体积即是面积，对 $n=3$，广义体积就是通常所说的体积．

据我所知，下面的问题迄今尚未解决：

14. 对哪一个正整数 n，装箱问题都有解？也就是说，对任意正数 $x_1,x_2,\cdots,x_n$，具有同样尺寸 $(x_1,x_2,\cdots,x_n)$ 的 n^n 个木块可以妥帖地装入一个棱长为 $x_1+x_2+\cdots+x_n$ 的完全 n 维立方体？

在 $n=1$ 时，问题是肤浅的．在 $n=2$ 或 $n=3$ 时，图 1 与图 2 已告诉我们，对任意的 (x_1,x_2) 或 (x_1,x_2,x_3)，应该如何着手．那么，对 $n=4$ 或更大的数，情况又怎样呢？

让我提醒读者，不等式 13 并不足以保证那些木块能够装入箱子，它只不过教我们别干那种毫无成功希望的蠢事．（如果 13 式得不到满足，我们就可以肯定，木块无法妥帖地装箱．）

让我们把装箱法存在的那种维数 n 称为优美维数. 由上所述，可知 1,2,3 都是优美维数. J·赛弗里奇(J. Selfridge)告诉我，R·罗宾孙已证明了下述定理：

15. 如果 m 与 n 是优美维数，则 mn 也是优美维数.

特别地说，由于 2 与 3 都是优美维数，所以 4,6,8,12，甚至 $181398528=2^{10}\cdot 3^{11}$ 全都是优美维数. 于是，$4^4=64$ 个四维木块，每个都具有尺寸(x_1,x_2,x_3,x_4)，将能妥帖地装入一只棱长为 $x_1+x_2+x_3+x_4$ 的四维完全立方体内.

人们可能会怀疑所有的正数维 n 都是优美维数，如果确系如此，那就只需对一切素数 n 来研究就行了，因为任何正数都可表为素数的连乘积，而解法也可由 15 式"相乘"而得. 在目前，最小的存疑维数是 5.

我希望某些满怀雄心的读者会去尝试求解如下问题：$5^5=3125$ 个五维木块(每个大小为 $x_1\times x_2\times x_3\times x_4\times x_5$)能否装入一个边长为 $x_1+x_2+x_3+x_4+x_5$ 的五维立方体箱子？

在理解罗宾孙对 15 式的证明时只有一个绊脚石，那就是下文将要说明的 16 式. 一旦你搞通了 16 式，你就会发现，证明 15 式是很容易的.

在我们叙述这个事实以前，让我们给出一个例子. 在图 1 中我们已表明，四只(7,8)矩形可以装填一个(15,15)正方形. 因此，四块(7,8,100)砖块将可装入一个(15,15,100)的箱子. 类似地，四只五维砖块(每块的大小都是(7,8,100,14,47))将能装入一只五维(15,15,100,14,47)箱子.

16. 如果其大小为$(x_{1,1},x_{1,2},\cdots,x_{1,n})$，$(x_{2,1},x_{2,2},\cdots,x_{2,n})$，$\cdots$，$(x_{k,1},x_{k,2},\cdots,x_{k,n})$的 k 只 n 维砖块，能妥帖地装入一只大小为$(y_1,y_2,\cdots,y_n)$的箱子. 则大小为$(x_{1,1},x_{1,2},\cdots,x_{1,n},z_1,z_2,\cdots,z_t)$，$(x_{2,1},x_{2,2},\cdots,x_{2,n},z_1,z_2,\cdots,z_t)$，$\cdots$，$(x_{k,1},x_{k,2},\cdots,x_{k,n},z_1,z_2,\cdots,z_t)$的 $k(n+t)$维砖块可以装进一只大小为$(y_1,y_2,\cdots,y_n,z_1,z_2,\cdots,z_t)$的箱子.

现在我们准备证明罗宾孙定理(15 式)，如果已知 m 与 n 是优美

维数，我们要证明 mn 也是一个优美维数. 换言之，我们必须找到一种办法，把具有一样大小 $(x_{1,1}, x_{1,2}, \cdots, x_{1,n}, x_{2,1}, x_{2,2}, \cdots, x_{2,n}, x_{3,1}, \cdots, x_{(m-1),n}, x_{m,1}, x_{m,2}, \cdots, x_{m,n})$ 的 $(mn)^{mn}$ 个 mn 维砖块装入一只 mn 维完全立方体箱子，此箱的各边之长均为 t，而 t 则为 mn 个尺寸 $x_{1,1}, \cdots, x_{m,n}$ 之和. 如果我们不把砖块的 mn 维尺寸都排成长长的一行，而改用下面的矩阵形式

$$\begin{bmatrix} x_{1,1}, & x_{1,2}, & \cdots, & x_{1,n} \\ x_{2,1}, & x_{2,2}, & \cdots, & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m,1}, & x_{m,2}, & \cdots, & x_{m,n} \end{bmatrix}$$

表达，则证明过程将会较清楚与简单一些.

第一步，我们把 $(mn)^{mn}$ 个砖块分成每群都有 m^m 块砖的各群. 显然群的个数应是 $(mn)^{mn}/m^m = m^{m(n-1)}n^{mn}$，令 $s_1 = x_{1,1} + x_{2,1} + \cdots + x_{m,1}$，既然 m 是一个优美维数，故知每块尺寸均为 $(x_{1,1}, x_{2,1}, \cdots, x_{m,1})$ 的 m^m 块砖将能装入一只尺寸为 $(s_1, s_1, \cdots, s_1)$ 的箱子. 利用 16 式，我们可以看出，$m^{m(n-1)}n^{mn}$ 群中的任意一群所有的 m^m 块砖能装入一个箱子，其尺寸为

$$\begin{bmatrix} s_1, & x_{1,2}, & \cdots, & x_{1,n} \\ s_1, & x_{2,2}, & \cdots, & x_{2,n} \\ \vdots & \vdots & & \vdots \\ s_1, & x_{m,2}, & \cdots, & x_{m,n} \end{bmatrix},$$

这类箱子有 $m^{m(n-1)}n^{mn}$ 只. 再把这些新的箱子分群，使每一群均有 m^m 只箱子，这时群的个数将是 $m^{m(n-1)}n^{mn}/m^m = m^{m(n-2)}n^{mn}$. 在每一群中的 m^m 只箱子可以装入一只尺寸为

$$\begin{bmatrix} s_1, & s_2, & x_{1,3}, & \cdots, & x_{1,n} \\ s_1, & s_2, & x_{2,3}, & \cdots, & x_{2,n} \\ \vdots & \vdots & \vdots & & \vdots \\ s_1, & s_2, & x_{m,3}, & \cdots, & x_{m,n} \end{bmatrix}$$

的箱子，此时 $s_2 = x_{1,2} + x_{2,2} + \cdots + x_{m,2}$.（这里我们利用了以下事实：每块大小均为 $(x_{1,2}, x_{2,2}, \cdots, x_{m,2})$ 的 m^n 块砖能装入一只 $(s_2, s_2, \cdots, s_2)$ 的箱子并再次应用 16 式.）

把每次过渡一列的过程重复进行 n 次. 在每一阶段，把上一阶段所得的箱子进行分群，使每群具有 m^n 只箱子，再把每群中的箱子重新集合为一只新的箱子. 这些新箱子的尺寸（每群一只）是把列中所有元素之和来取代正被谈论的列的每个元素. 其结果是：$(mn)^{mn}$ 个原始砖块都被装进 $m^{m(n-n)} n^{mn} = n^{mn}$ 只箱子，其中每只箱子的各个尺寸为

$$\begin{pmatrix} s_1, & s_2, & \cdots, & s_n \\ s_1, & s_2, & \cdots, & s_n \\ \vdots & \vdots & & \vdots \\ s_1, & s_2, & \cdots, & s_n \end{pmatrix},$$

这里，$s_i = x_{1,i} + x_{2,i} + \cdots + x_{m,i}$；$i = 1, 2, \cdots, n$.

现在我们来做同样的事情，不过，这时是对行而不是对列了. 把 n^{mn} 只箱子分成每群具有 n^n 只箱子的各群. 群的个数将是 $n^{mn}/n^n = n^{(m-1)n}$，既然 n 是一个优美维数，所以，大小为 $(s_1, s_2, \cdots, s_n)$ 的 n^n 个砖块能装入一只大小为 $(t_1, t, \cdots, t)$ 的箱子，这里 $t = s_1 + s_2 + \cdots + s_n$（注意 t 是所有 mn 块原始砖块的尺寸 $x_{1,1}, x_{1,2}, \cdots, x_{m,n}$ 之和，这在 16 式的下面已经定义过，因此 t 是我们企图装入的 mn 维超立方体的每条边之长）. 于是，利用 16 式，我们可以看出，每一群中的 n^n 只箱子将能装入一只大小为

$$\begin{pmatrix} t, & t, & \cdots, & t \\ s_1, & s_2, & \cdots, & s_n \\ \vdots & \vdots & & \vdots \\ s_1, & s_2, & \cdots, & s_n \end{pmatrix}$$

的新箱子，而这种箱子共有 $n^{(m-1)n}$ 只. 对 m 行的每一行都重复以上过程，其结果是一切砖块都能装进一只 $(n^{(m-m)n} = 1)$ 大小为

$$\begin{pmatrix} t, & t, & \cdots, & t \\ t, & t, & \cdots, & t \\ \vdots & \vdots & & \vdots \\ t, & t, & \cdots, & t \end{pmatrix}$$

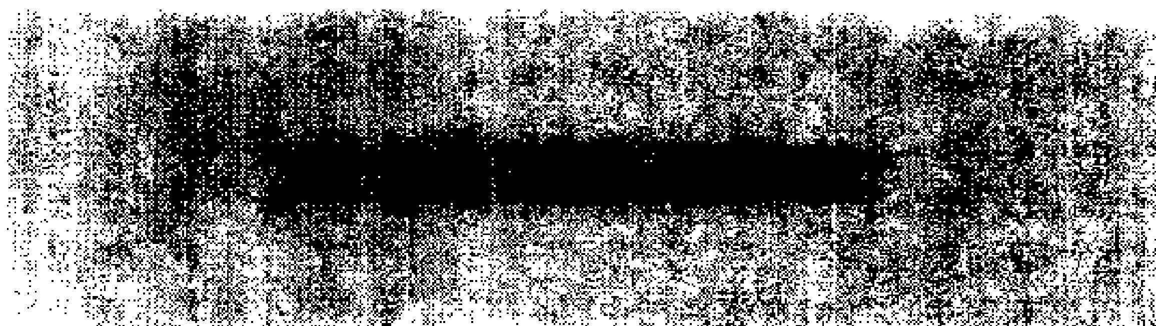
的箱子，而这正是我们的意图. 于是，所有$(mn)^m$只原始砖块都能装箱，也就是说，mn 是一个优美维数. 证明完毕.

在本章中，我们已经看到一些例子，藉以表明不等式与装箱问题的巧妙联系：有时候，已知不等式可以导出有趣的装箱问题；反之，有时候，可以通过求解一个装箱问题来证明一个不等式.

上述证明过程表明一种进一步联系. 在关于算术平均数与几何平均数不等式的某些标准证法中，一个步骤是要设法去证：若对 m 个数成立，又对 n 个数亦成立，则对 mn 个数同样成立. 而这种证法可以表述为，证明的每一步对应于上述罗宾孙证法中的一步！我们所获得的教益是：一个不等式的已知证法有时可用于找出一个装箱问题的解.

数学里头有很多不等式，它们中间肯定有许多可以导出有趣的装箱问题. 我想为读者推荐一本很优秀的读物《几何不等式》，作者是 N·D·卡查林诺夫(N. D. Kazarinoff)(Random House 出版的《新数学文库》第四册). 任何一个具有高中数学水平与虚心好学的人都能看得懂这本书(实际上，《新数学文库》中所有其他书籍也是如此).

当你下次遇到一个不等式时，请你想尽一切办法把它转变为一个装箱问题，也许你将为所导出的结果深深陶醉！



● 加利福尼亚大学

拉斐尔·M·罗宾孙(Raphael M. Robinson)

本文中,我们将研究由全等的立方体组成的空间铺砌.在说到这些立方体组成空间铺砌的时候,我们指的是它们在空间不重叠.我们设想这些立方体的棱都是平行于坐标轴的.由于运用单位立方体不至于损害一般性,下文我们将要经常采取这种做法.

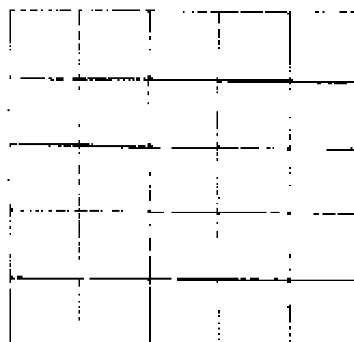


图 1

平面的标准铺砌.

我们首先从二维(较简单)情况着手研究,这里,我们所涉及到的是由边平行于坐标轴的全等正方形组成的平面铺砌.最简单的平面铺砌是如图1所示的"标准铺砌".在这平面铺砌中,每一个正方形都与其他四个正方形以边边相接的形式连接.仅当两个正方形拥有一

条公共边时我们说它们是边边相接. 我们也能够在标准平面铺砌中, 把不同的行移动不同的长度, 形成另外的平面铺砌, 如图 2 所示. 在这种情况下, 我们就组成这样的平面铺砌, 它里面的正方形只与其他两个正方形边边相接. 除了移动各行的位置, 我们也能移动各列的位置, 如图 3 所示. 从这里我们容易看到, 在任何一个由全等正方形组成的平面铺砌中, 每个正方形总是位于某一行或某一列, 或同时位于某行某列之中. 因此, 这类铺砌中, 每一个正方形至少与其他两个正方形边边相接. (平面或空间中的格点是指坐标为整数的点 (x,y) 或 (x,y,z).) 如果我们把一个单位正方形的中心放在平面的每一个格点上, 那么我们将得到一个标准的平面铺砌.

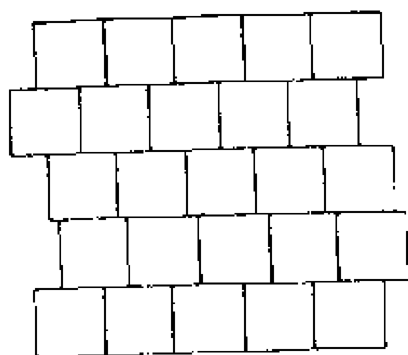


图 2
移动行以后的平面铺砌.

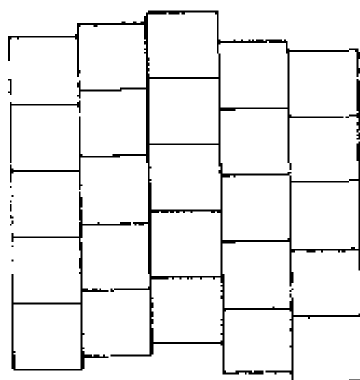


图 3
移动列以后的平面铺砌.

现在, 我们转向三维情形. 如果我们把一个单位立方体的中心放在空间的每一个格点上, 那我们就获得了一个标准的空间铺砌. 在这类空间铺砌中, 每个立方体都与其他六个正方体以面面相接的形式连接(我们所说的面面相接仅仅指它们拥有一个公共面的状态). 这类空间铺砌可以分解成平行于 xy 平面的不同层. 要改变这类空间铺砌, 可以在 x 及 y 方向任意移动那些不同的层. 每一层之中, 我们可以把行在 x 方向移动任意长度, 或者把列在 y 方向移动任意长度, 就如在平面铺砌中一样. 我们也可以在某些层中移动行, 在另一些层中

移动列.通过这类移动,我们能得到这样的空间铺砌:它的每个立方体仅同其他两个立方体面面相接.

从平行于任一坐标平面的层出发,我们亦可以得到类似的结构.但是,每一个立方体并不一定要落在某一层中;事实上,我们可以给出几个不存在层的空间铺砌.在说明这个问题时,为了方便起见,我们用"列"这个词来代表那些在任一坐标方向连接起来的一串立方体.

在标准铺砌的三个坐标方向上各挑选一列,使它们不具有公共立方体,然后沿各自坐标轴方向移动,我们便得到了一个最简单的没有层的空间铺砌.

这一概念的推广将能得出使某一特定的立方体不与任何其他立方体面面相接的空间铺砌.为了组成这样的空间铺砌,我们先从标准铺砌开始,挑选一个准备避免与其他任何立方体面面相接的立方体.它周围的六个立方体可被归入六列——每个坐标方向上两列——而每一列都能自由移动,不牵连其他列或中心立方体.例如,在 x 方向上邻近中心立方体的立方体可以在 y 方向上移动;在 y 方向邻近中心立方体的立方体可以在 z 方向移动;在 z 方向邻近中心立方体的立方体可以在 x 方向移动.每一种情况,都是在指明的坐标方向上作整列移动.这样中心立方体就不再与其他立方体面面相接了.

与此类似,我们可以作出一个空间铺砌,它里面具有无限多个避免与其他立方体面面相接的立方体.下文所描述的空间铺砌是由汉斯·强生(Hans Jansen)于 1909 年发现的.首先把每个立方体的中心放在空间每一个格点 (x,y,z) 上形成一个标准的空间铺砌.然后将以下立方体的中心 (x,y,z) 移动 $\frac{1}{2}$ 个单位:

如果 y 为偶数,z 为奇数,则沿 x 方向移动;

如果 z 为偶数,x 为奇数,则沿 y 方向移动;

如果 x 为偶数,y 为奇数,则沿 z 方向移动.

当 c 为偶数或奇数时,截面 $z=c$ 的结构如图 4 所示.这些平面通过

许多立方体的中心，同时，截面也通过了某些沿 z 方向移动的立方体的表面，如图 4 中阴影部分所示. 箭头表示相对其他立方体移动的方向. 那些 x、y、z 坐标全为奇数或全为偶数的立方体是固定的，因而没有被标记. 它们不与其他立方体面面相接. 显然，这些不与其他立方体面面相接的立方体占全部的 $\frac{1}{4}$. 另一方面，如果给定一个由全等立方体组成的空间铺砌，不难发现，假如其中一个立方体的一面没有与其他立方体面面相接，那么一定有两个与此面邻接的立方体自己面面相接. 因此，仅仅是某些立方体能够避免面面相接，而不是所有的.

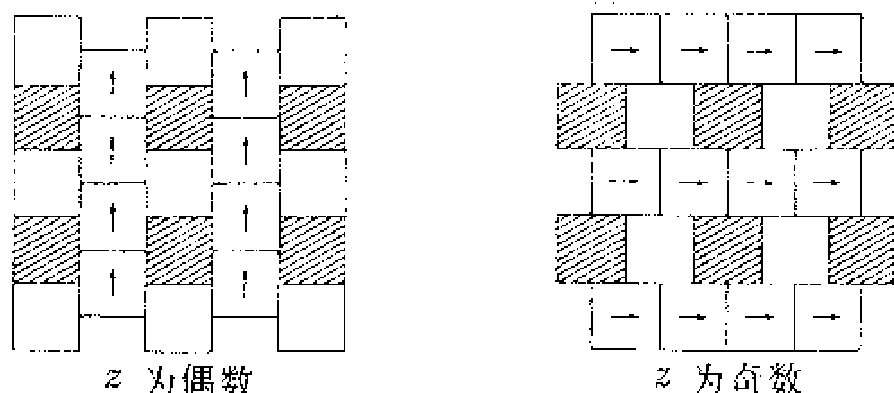


图 4

空间的强生(Jansen)铺砌的截面.

我们现在来考虑由全等正方形或全等立方体组成的多重铺砌. 我们假设平面或空间的每个点被覆盖有限次，而且那些不在正方形或立方体边缘上的点被覆盖同样的次数. 如果这数是 k，那么我们称之为 k 重铺砌. 我们先前考虑的铺砌都是 1 重铺砌，或称简单铺砌. 平面多重铺砌并不是很有趣的. 不难看出，任何一个 k 重平面铺砌都可以由 k 个简单铺砌叠成. 因此，每个正方形至少与其他两个正方形边边相接.

在空间，出现了些新情况，并不是所有的多重空间铺砌都可以由简单铺砌叠成. 进一步说，使每一个正方体避免面面相接是有可能的. 这在 1974 年由我证明，当时我发现了全等立方体叠成的 25 重空

间铺砌,其中任何两个立方体都不面面相接.在描述这类空间铺砌时,为了方便起见,我们用边长为 5 的立方体代替单位立方体.显然,这种尺寸上的替换不会影响问题的结果.我们所用到的立方体的中心都位于格点上.我们将利用满足下列四个条件之一的中心点 (x,y,z):

$$x\equiv y\equiv z(\mathrm{mod}2),\qquad x+y+z\equiv 0(\mathrm{mod}5);$$
$$x+1\equiv y\equiv z(\mathrm{mod}2),\quad x+y+z\equiv 1(\mathrm{mod}5);$$
$$x\equiv y+1\equiv z(\mathrm{mod}2),\quad x+y+z\equiv 2(\mathrm{mod}5);$$
$$x\equiv y\equiv z+1(\mathrm{mod}2),\quad x+y+z\equiv 3(\mathrm{mod}5).$$

这里我们所使用的同余式记号 $a\equiv b(\mathrm{mod}\ m)$ 是指 a 与 b 的差为 m 的倍数.上面的第一行指的是坐标 x、y、z 同是奇数或者同是偶数,并且它们的和是 5 的倍数.总之,每一行的第一部分总是说明三个坐标是奇数还是偶数;而第二部分则给出了它们的和除以 5 的余数.第一个条件称为奇偶性条件.所有满足上述某一奇偶性条件的格点全体被称为组成一个奇偶类.我们将证明所有上面提到的立方体加在一起能组成一个 25 重空间铺砌,而且其中没有任何两个立方体面面相接.

为了说明它是一个 25 重空间铺砌,就有必要证实每一个格点都被覆盖 25 次.现在,每一个以 5 为边长,包含格点 (x',y',z') 的立方体的中心点集正好是以 (x',y',z') 为中心,边长为 5 的立方体内的各个格点 (x,y,z).首先,我们考虑 $(x',y',z')=(0,0,0)$ 的情形,也就是一个中心在原点、边长为 5 的立方体.这个立方体内的格点 (x,y,z) 共有 125 个,对应坐标值分别为 $-2,-1,0,1,2$.这就形成了一个 $5\times5\times5$ 的立方阵.我们先来看满足条件 $x\equiv y\equiv z(\mathrm{mod}2)$ 的那个奇偶类.它包含两个部分,当 x、y、z 均为偶数时,形成 $3\times3\times3$ 立方阵,共 27 点;当 x、y、z 均为奇数时,形成 $2\times2\times2$ 立方阵,共 8 点.

现在让我们构作截面 $x+y+z=c$,它正好垂直于立方体的一条对角线.$3\times3\times3$ 立方阵的截面将含有 1,3,6,7,6,3,1 个点,如图 5 所示.$2\times2\times2$ 立方阵的截面将含有 1,3,3,1 个点.把这 35 个点加

在一起考虑，我们发现它们所在的等间隔的平面包含 1,0,3,1,6,3,7,3,6,1,3,0,1个点. 把 $x+y+z \pmod 5$ 取同值的点结合起来，将可得到如下总和：

	1	0	3	1
6	3	7	3	6
1	3	0	1	
7	7	7	7	7

因此，在给定的立方体中正好有 7 个格点满足 $x \equiv y \equiv z \pmod 2$，并符合 $x+y+z \pmod 5$ 的预定值.

给定立方体的所有格点组成上面所提到的 $5\times5\times5$ 立方阵，我们再一次考虑截面 $x+y+z=c$，其中的三个已在图 6 中给出. 截面中所含点的个数列在下面；它们是按 $x+y+z \pmod 5$ 的值组合的.

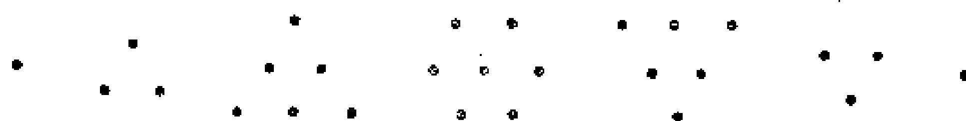


图 5

$3\times3\times3$ 立方阵的斜截面.

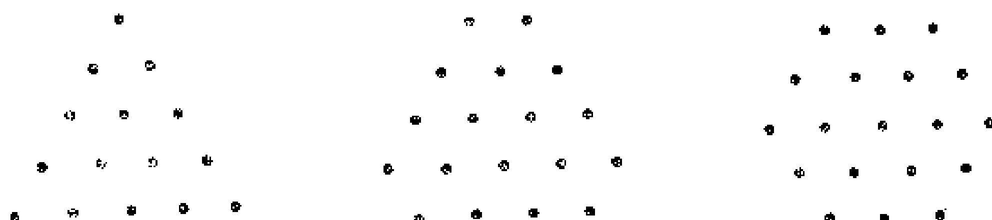


图 6

$5\times5\times5$ 立方阵的斜截面.

	1	3	6	10
15	18	19	18	15
10	6	3	1	
25	25	25	25	25

可见每个值都出现了 25 次. 这件事也可以从另一角度来证明，当 x、

y 取 $-2,-1,0,1,2$ 中任意值时，则对边给定了的 $x+y+z \pmod 5$，z 是唯一确定的. 满足 $x\equiv y\equiv z \pmod 2$ 的每个 $x+y+z \pmod 5$ 的值出现 7 次，因而对其余的点出现 18 次. 现在我们把满足其他 3 个条件的奇偶类以 x、y、z 进行轮换. 换句话说，其他 3 组奇偶类的格点是以对空间直线 $x=y=z$ 相继旋转 120°的方式进行置换.

因此，在每一组中，任何 $x+y+z \pmod 5$ 的指定值在给定的立方体中必然正好出现 6 次.

现在让我们考虑中心在任意点 (x',y',z')、边长为 5 的立方体. 如果我们把这个点平移到原点，那么在给定立方体中的格点 $(x、y、z)$ 将变为格点 $(x-x',y-y',z-z')$，如果 (x',y',z') 是给定的，那么这点的奇偶类就由 $(x、y、z)$ 的奇偶类来确定，反之亦然. 特别地，如果 $(x、y、z)$ 正落在 (x',y',z') 的同一奇偶类时，我们将有：

$$x-x'\equiv y-y'\equiv z-z' \pmod 2).$$

从上面的论证结果来看，给定的立方体中存在7个点满足给定的 $x+y+z \pmod 5$ 的值. 但在其余三组中，都仅有 6 个点满足给定的 $x+y+z \pmod 5$ 的值.

由此可见，在空间铺砌中那些满足四个条件之一的立方体将覆盖满足同样奇偶性条件的所有格点 7 次，而覆盖其余格点 6 次. 那么，四组立方体加在一起便覆盖所有的格点 25 次，因而就形成了一个 25 重的空间铺砌. 所有这一切并不受四个条件中 $x+y+z \pmod 5$ 之值的约束.

我们现在验证没有两个立方体以面面相接. 如果有两个立方体面面相接，那么第二个立方体的中心一定能通过移动第一个立方体的中心（使其中的一个坐标改变 5 个单位长）而得出. 这样就改变了奇偶类，然而却没有改变 $x+y+z \pmod 5$ 的值——然而这种事情是不可能发生的，因为四个奇偶类中 $x+y+z \pmod 5$ 的值是不等的. 这样我们就得到了一个没有两个立方体面面相接的 25 重空间铺砌.

请注意，这个由边长为 5 的立方体组成的空间铺砌，沿任意坐标轴方向移动 10 个单位并不会改变整个空间铺砌，或者按向量(5,5,

5)进行平移也是一样.可以证明,这些性质是没有两个立方体面面相接的必然推论.

要想通过画图来说明这种空间铺砌是极其困难的.我们所能采取的最好办法无非是画截面,并标出立方体的中心.

在图 7 中,我们给出两个截面,即 $z=0$ 和 $z=25$. 这里画出中心位于格点 (x,y,z) 的单位立方体截面,其中 x 和 y 由 1 至 9 变化.图中的点标记着这一 25 重空间铺砌中边长为 5 的立方体的中心.这两张图无论在水平或垂直方向都反复出现.其他截面也有与此相似的外形.

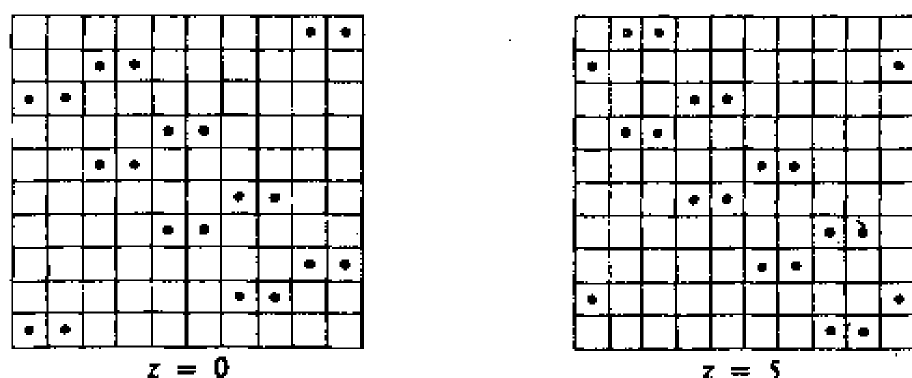


图 7

25 重空间铺砌的截面.

为了获得空间的多重铺砌,必须满足:处于同样位置的两个截面(5×5 正方形)一定含有同样的点数.为了使任意两个立方体避免面面相接,那么就该使得点在两个截面上处于不同的位置,并且两个面上没有任何两点在水平或垂直方向上相距 5 个单位.

最近在文献〔1,第 12 节〕上发表了与上文所述相同的 25 重空间铺砌,但其证法略有不同(在〔1〕中主要讨论高维情况下相应的问题,但附加了一个条件:立方体的中心组成一个点阵.)至于其他有关的结果和史料,请参看〔2〕.

当 $k<25$ 时,是否存在着由全等立方体组成的、完全避免面面相

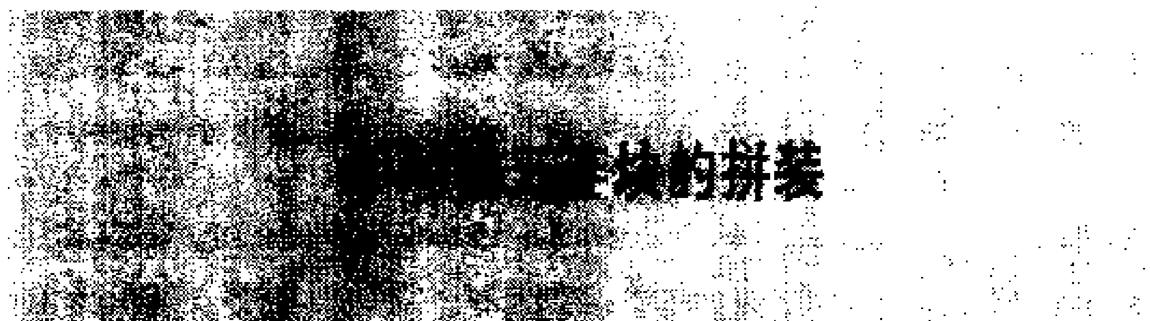
接的空间铺砌? 依然悬而未决.

参考文献

1 Robinson, R. M. 1979. Multiple tilings of n-dimensional space by unit cubes. *Mathematische Zeitschrift* 166 : 225—264.

2 Stein, S. K. 1974. Algebraic tiling. *American Mathematical Monthly* 81 : 445—462.

附言:此文写好以后,巴赛尔·戈登(Basil Gordon)告诉我,$k=2$时也是可能的.实际上,存在由边长为2的立方体组成的空间铺砌,每一个由八个单位正方体组成标准铺砌,其中没有两个立方体是面面相接的.



● 埃因侯温工艺技术学校

□ C·J·博坎普(C. J. Bouwkamp)

读者也许还记得，一共有 29 种形状各不相同的立体五连块. 所谓五连块，就是把五个单位立方体连在一起的立体，其中每两个相邻的立方体都是以整个面相接合的，并且每个立方体至少有一个与之相邻的立方体. 这 29 个五连块中，有在数学游戏领域中为人们所熟悉的 12 个平面五连块，也被称作“潘多米诺”(pentomino). 把这 12 块加在一起，体积一共是 60 个单位，可以装入尺寸大小各不相同的箱子，例如 $2\times3\times10$，$2\times5\times6$，$3\times4\times5$ 等等(可参阅文献[1,2,3]). 除了以上 12 块外，剩余的五连块中还有五个是具有至少一个对称平面的，它们中每一个都是自己本身的镜像. 再剩余下来的 12 个五连块组成了六对，每对包括两个互为镜像的五连块立体. 每对五连块就像一个人的左手同右手的关系，我们称之为“手对称的五连块”. 图 1 中展示了所有成对的手对称的五连块，并以 1—12 的数字作为它们的标号. 与平面五连块相同的是：手对称的五连块加在一起的总体积也是 60 个单位，这就使人们很自然地要问：这种新 12 块是否也能装入类似尺寸为 $3\times4\times5$ 之类的盒子呢?

澳大利亚新南威尔士州立大学化学系的 G·V·巴德莱(G. V. Baddeley)博士曾于 1973 年 7 月通过信件向我提出了这个问题. 他主要是研究立体化学的，他在信中写道：“是否能用手对称的十二块

装进一只大小为 3×4×5 的箱子，如果能够的话，那么答案是否唯一？”

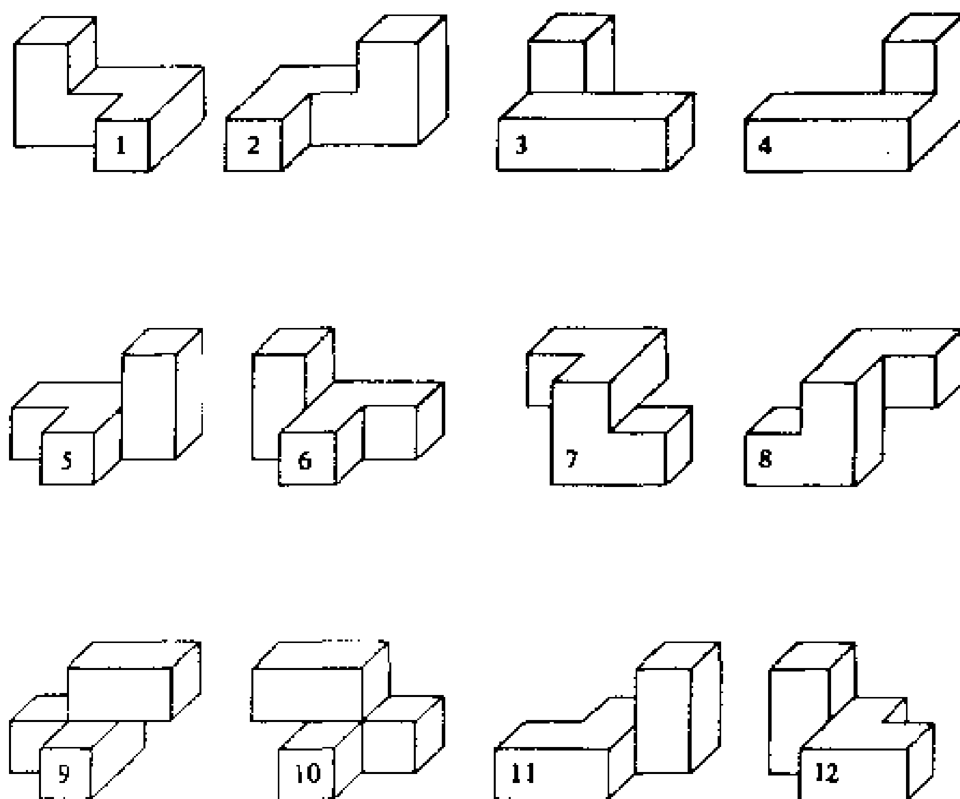


图 1

手对称五连块及其标号.

我插一句，如果读者有幸能搞到一套手对称的十二块，我建议你在阅读下文以前自己动手寻找答案. 虽然用它们装进一只盒子比用平面十二块困难得多，但你也不必马上放弃，因为现在已经可以证明不同的装法为数甚多.

埃因侯温菲利浦研究实验室的 J·M·M· 韦勃凯尔(J. M. M. Verbakel)发现了一种既灵活简便又同时能得到多种拼法的方式. 他把整套手对称的十二块分成四组，每组三块，以便能同时拼成四块，这里我们称作“钢琴”. 图 2 中展示了一个韦勃凯尔氏结构，其中左边两个“钢琴”是右边两个“钢琴”的镜像(对称图形). 显然，四个“钢琴”可以同时装进 2 个 2×3×5 盒子(有 3 种不同的拼法)，因此可装进

一个 3×4×5 盒子，那就解决了巴德莱博士提出的存在性问题．而且，拼法根本不是唯一的．作为韦塞凯尔氏结构的副产品，我们可以看出用

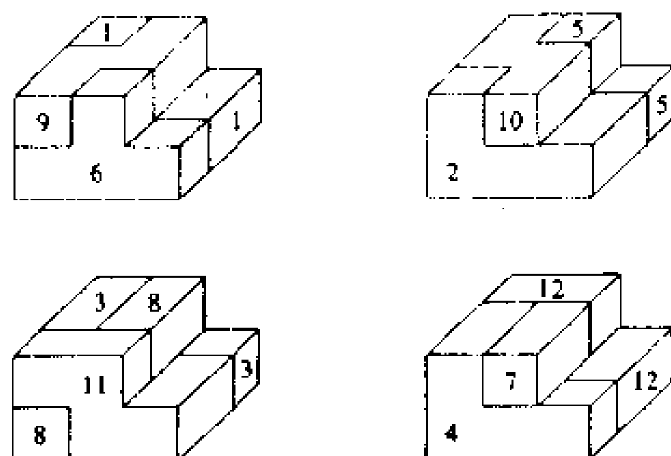


图 2

由手对称五连块同时构成的四个钢琴．

1	(1,6,9)	(2,5,10)	(3,8,11)	(4,7,12)
2	(1,6,9)	(2,5,10)	(3,8,12)	(4,7,11)
3	(1,7,9)	(2,8,10)	(3,5,11)	(4,6,12)
4	(1,7,9)	(2,8,10)	(3,6,12)	(4,5,11)
5	(1,6,9)	(2,8,10)	(3,5,11)	(4,7,12)
6	(1,7,9)	(2,5,10)	(3,8,11)	(4,6,12)
7	(1,8,11)	(2,9,12)	(3,4,7)	(5,6,10)
8	(1,10,11)	(2,7,12)	(3,4,8)	(5,6,9)

表 1

构成同时存在的四个“钢琴”的手对称五连块的 8 种划分办法．

这十二块同样可以组成 $2\times3\times10$ 立体和 $2\times5\times6$ 立体，当然也不止一种答案(与平面十二块一样).

利用计算机，我找到了用手对称十二块划分成四个“钢琴”的所有方法，其中有 8 种在旋转和平移下实质不同的办法(在反射、旋转和平移下实质不同的办法有 6 种)，这 8 种划分办法列在表 1. 前面的 4 种划分方法应归功于李勃凯尔，它们是自身对偶的. 划分方法 5 和 6 是相互对偶的. 同样 7 和 8 也是如此. (请读者自己动手将表 1 中所列出的其他 7 种划分拼出来(事实上这是摆弄手对称十二块的一个美妙的练习). 你会发现，将有两个不同的(5,6,9)钢琴，和两个不同的(5,6,10)钢琴.)

现在，我该考虑如何复合成 $3\times4\times5$ 的立方块了. 我们知道，立方块是由尺寸较小的长方体拼成. (顺便说一下，$3\times4\times5$ 立方块必定可由两个 $2\times3\times5$ 较小的立方块组成.)利用计算机，我确定了所有的复合成 $3\times4\times5$ 的立方块的组成部分，其结果列于表 2 之中. 根据拼成第一个立方块(均包含标号为 1 的五连块)的六个不同的五连块，我区分了 24 种类型. 表中列出了用给定六个五连块拼成不同的 $2\times3\times5$ 立方块的方法数，以及没有指出来的六个五连块拼第二个立方块的方法数. 每种类型的不同的组合数列在最后一列中(包括经反射，旋转，还有平移等变换后实质不变的方法). 其中字母 S 代表“自身对偶”. 这些数加在一起为 $28+30S$，它的意思是同时组成两个 $2\times3\times5$ 立方块有(对旋转来说)$2\times28+30=86$ 种不同的组合. 顺便说一下，只是一些立方块而不是所有的立方块都是由“钢琴”拼成的，例如图 3 展示了类型 3 立方块可有 3 种方法拼成第一块，有 4 种方法拼成第二块. 通常，每块由两个矩形截面来表示，其中之一为底面(左图)，另一为顶面(右图). 我们可以清楚地看到，七块中只有三块是由“钢琴”拼成的. 值得注意的是，所有这七块都是对称的，也就是说，每块都有一个对称中心.

组成 $3\times4\times5$ 立方块的不同方法(对旋转与映射而言)共有 37 种. 其中，28 种是有一个对称平面的，其余更多的则有一个对称中

类型	第一块中的五连块	解的个数		共计（S 代表自身对偶）
		第一块	第二块	
1	(1,2,5,6,9,10)	1	3	3S
2	(1,2,5,7,9,10)	1	1	1
3	(1,2,7,8,9,10)	3	4	12S
4	(1,2,7,8,11,12)	1	1	1S
5	(1,2,7,10,11,12)	1	2	2
6	(1,3,4,5,9,11)	3	1	3
7	(1,3,4,6,7,10)	1	1	1
8	(1,3,4,6,8,9)	1	2	2
9	(1,3,4,7,8,11)	1	2	2
10	(1,3,5,6,8,10)	1	1	1
11	(1,3,5,6,9,11)	4	1	4
12	(1,3,5,7,9,11)	2	2	1+2S
13	(1,3,5,8,10,12)	1	1	1S
14	(1,3,6,7,9,12)	1	1	1S
15	(1,3,6,7,10,12)	2	2	1+2S
16	(1,3,6,8,9,11)	1	1	1S
17	(1,3,6,8,9,12)	2	2	1+2S
18	(1,3,6,8,11,12)	2	1	2
19	(1,4,5,7,9,11)	1	1	1S
20	(1,4,6,7,9,10)	1	1	1
21	(1,4,6,7,9,11)	1	1	1S
22	(1,4,6,7,9,12)	3	3	3+3S
23	(1,5,6,8,9,10)	1	1	1
24	(1,5,6,8,10,11)	2	[illegible]	2
				共计 28+30S

表 2

组成 3×4×5 立方块的 24 种类型.

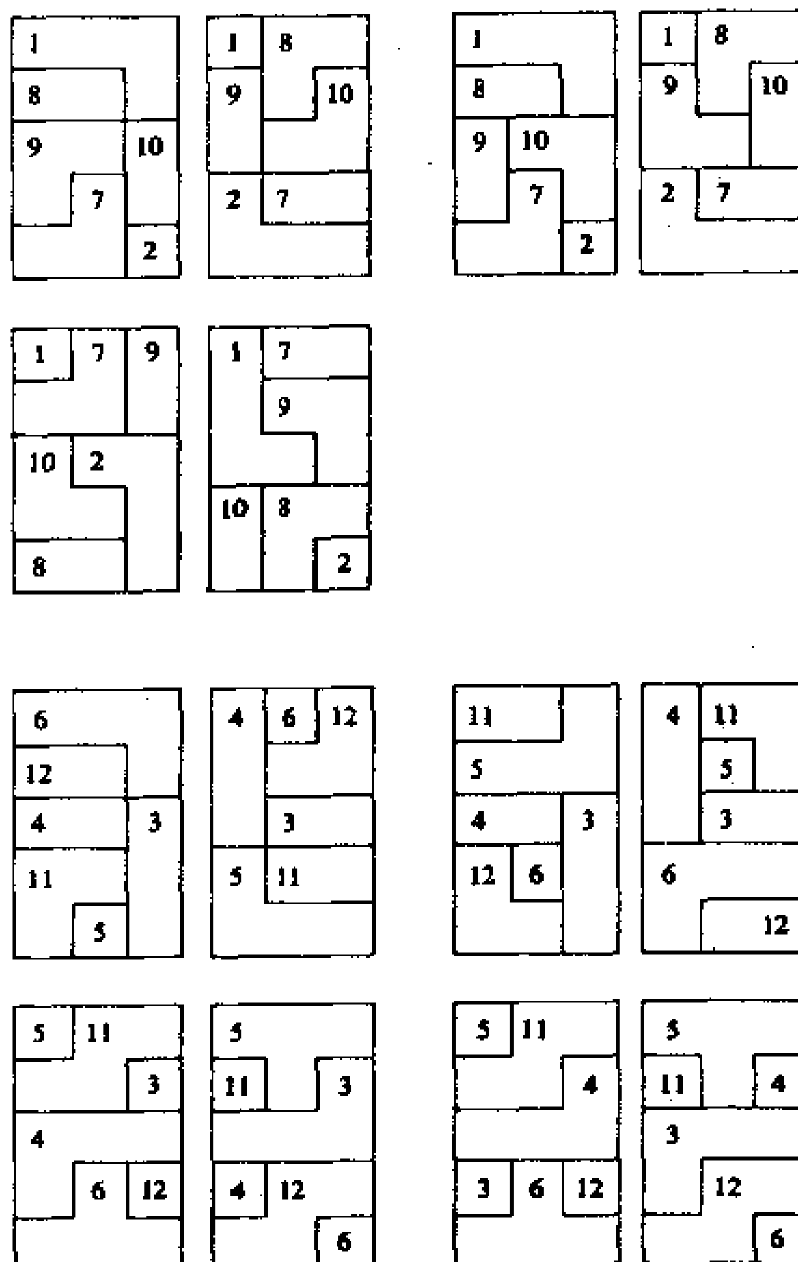


图 3

组成 3×4×5 立方块的类型 3 的基础块 2×3×5.

心. 有对称平面的 28 种组合是完整的一套,因为有对称平面的 3×4×5 立体必然是叠合的. 对于这些对称的叠合立体的实际结构,我建议读者看看图 4 所展示的 2×3×5 立方块,它是由标号为(1,3,5,8,10,12)的五连块拼成的. 与之对应的"对偶"块可以由剩余的标号为

(2,4,6,7,9,11)的五连块拼成.这两块可以拼成一个有对称平面的 3×4×5 立体(有两种方法).它们同样也可以被组合成有对称中心的立体,也有两种方法.许多既是平面对称又是中心对称的 3×4×5 立体的其他例子可以通过图 2 所示的"钢琴"划分而发现.

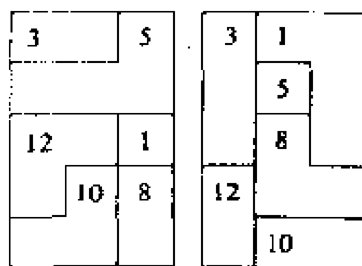


图 4

这个立体和它对偶图形能被装配成既是中心对称也是平面对称的 3×4×5 的立体,而且各有两种不同的装配法.

现在,我将转入讨论单层的 3×4×5 块,也就是非叠合的.证明这种方法的存在的一个简单方法是参考图 5 所示,用四个同时存在的"钢琴"装配组成一个单层的立体的结构.由于图 5 中的"钢琴"可以有几种变换存在,而又有 8 种不同的方法拼成一个"钢琴"(见表 1),显然,这种单层的拼块有许多种.

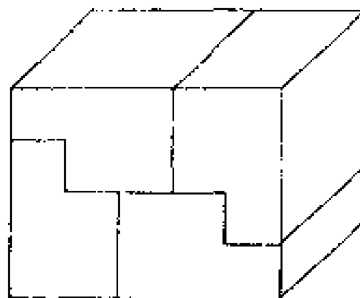


图 5

四个同时存在的钢琴装配成 3×4×5 的单层立体.

我曾试图寻找一种"信手得来"的方法,但总是不成功.因此,前文中要求读者自行寻找解决方法是不公平的.不过,我认为现在已知

的几百种解答方案仅仅是所有解答方案的一小部分，我对此确信不疑.

把问题编成程序，利用计算机求得一种或几种解答方案是一回

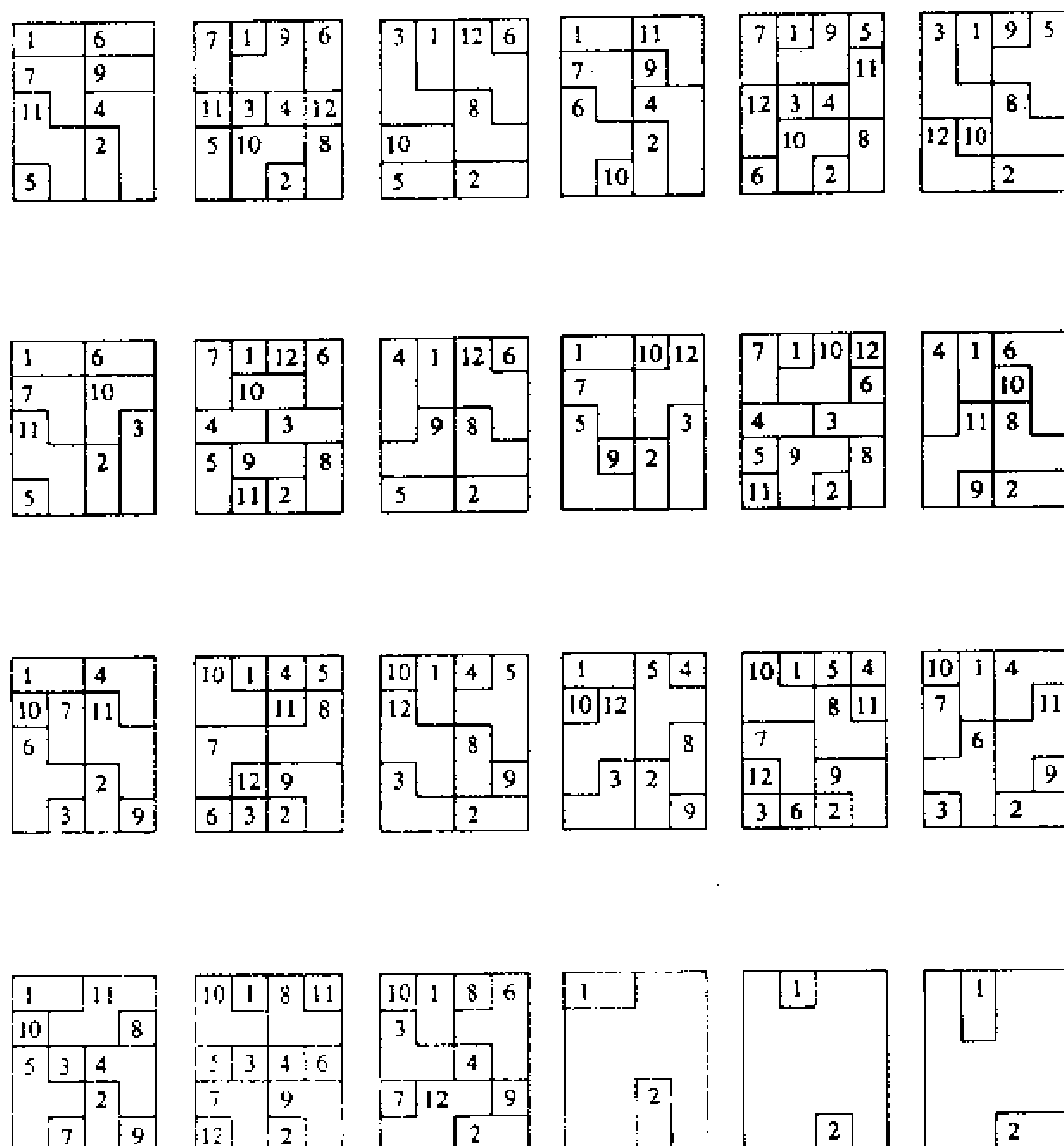


图 8

五个单层、两个叠合的 3×4×5 块，其中每个都具有对称中心．当 1,2 按图中位置固定时，此集合业已完备.

事，而求得全部的解答方案又完全是另外一回事．要消除反射和旋转是一件不容易的事，而解答方案中所反映出来的对称关系应在计算机的输出中有所表示．无论用一般性的程序还是用特殊的程序，我都可以发现拼成中心对称 3×4×5 立体的方法共有 742 种．这个数字包含了前面提到的 28 种叠合的立体．值得注意的是，它们只占总数不到百分之四的比例．

图 6 所展示的是标号为 1 和标号为 2 的五连块位于右下角所示位置时所有可能的解答方案，有一些是由上、中、下（或从左到右）三层叠合而成的．拼块中有五个是单层的，有两个是叠合的（类型 15）．

图 7 展示了有对称中心的美丽的拼块，它由两个阶梯组成，这两个阶梯可以装配成各种各样的立体．这类结构的立体有许多．

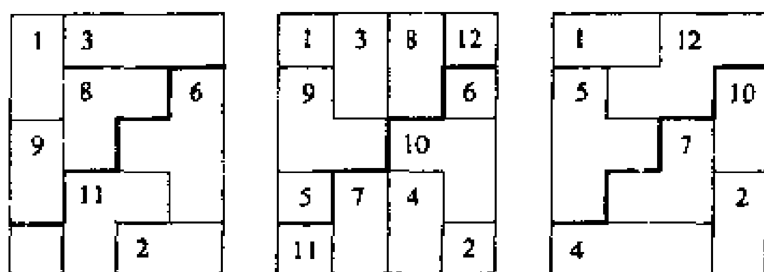


图 7

有一个对称中心的拼块可拆成两个"阶梯"，可以用两个"阶梯"以各种方式重新装配成 3×4×5 拼块．

最后给出的具有对称中心的拼块，见图 8 所示．它具有许多独特的性质．正如读者所观察到的一样，（7，8，9，10）四个五连块在拼块的三层图中没有出现．事实上，有两种可能填入所缺的标号，如图 8 下面五连块（7、8），（9、10）拼成的两个全等的立体．这样所装成的拼块是互为"映像"的（对旋转同态）．这种类型的结构非常少见，仅存在 10 种．

在本文即将结束之际，我们将要告诉大家用手对称十二块装配成 3×4×5 拼块的总方法数．无论你是否相信，对反射和旋转而言不

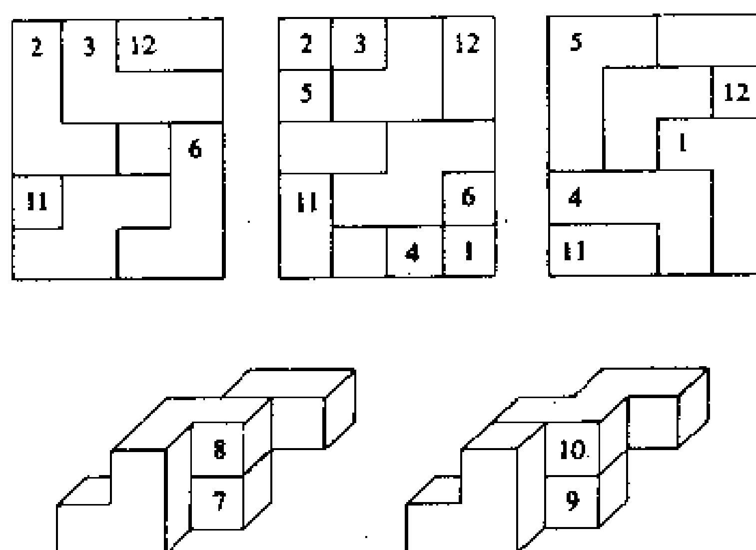


图 8

一个拼块可以通过平移 12 块五连块中的四块，形成它的“映像”的例子. 详见正文.

同的方法总数为 29162 种，对旋转而言，不同的方法为 57554 种. 这是利用计算机的空闲时间足足花了大约两年时间才获得的.

参考文献

1 Bouwkamp, C. J. *Catalogue of solutions of the rectangular $3\times4\times5$ solid pentomino problem*. 1967. The Netherlands: Technische Hogeschool Eindhoven, Department of Mathematics, Eindhoven.

2 ______ Packing a rectangular box with the twelve solid pentominoes. 1969. *J. Combinatorial Theory* 7: 278–280.

3 ______. Catalogue of solutions of the rectangular $2\times5\times6$ solid pentomino problem. 1978. *Kon. Ned. Akad. Wetensch., Proc., ser* A 81: 177–186.

我与多连骨牌打交道的经历

● 纽约州立大学

〔美〕戴维·A·克拉纳(David A. Klarner)

我从事数学研究工作也许是受了我父亲的影响，他有一个科学的头脑，是个开拓型的人，认为自己能干任何事情. 我在念高中时又遇到一位好老师内摩·代布莱(Nemo Debely)，是他培养了我对科学和数学的兴趣. 他尤其鼓励我阅读在五十年代后期开始登在《科学美国人》杂志上的马丁·加德纳的专栏文章. 在我父亲的熏陶下，我也认为我能征服数学领域中的任何问题. 每当代布莱给我一些经典的几何题，数论题，或者组合数学之类的问题时，我总是满怀信心地着手工作，当然，我没有想去征服费马大定理，"四色猜想"，以及孪生素数的无限性等问题. 但在不断地尝试过程中，我确实获益匪浅. 就在这天真的年岁里，我初次与马丁·加德纳有了交往.

当马丁·加德纳给我写回信，并把我介绍给其他一些与我有共同志趣爱好的人时，你可以想象，作为一名高中生的我是多么兴奋.(例如，正是通过马丁，我第一次认识了所罗门·果隆姆(Solomon Golomb)，在他后来的著作《多连骨牌》中编入了许多我所研究的问题和结果.)

在我逐渐进入数学世界的过程中，我遇到了一系列杰出的导师：高中时的内摩·代布莱，大学时的詹姆士·E·豪斯霍尔德(James E. Householder)，大学研究所的李奥·莫塞(Leo Moser)，博士后时期

的狄克·德·布鲁因(Dick de Bruijn).当然,马丁·加德纳也属于这一系列之中,不过他的作用和其他人又不一样.二十年来,他对我来说始终是一位鼓舞者,他让我与许多人分享数学的乐趣.

在本文中,我试图陈述一些我在多连骨牌方面的数学发现——当然对通俗解释来说这个课题似乎技术性太强了.虽然如此,我希望写出的东西使外行的人也能看懂.本文的前面几节讨论计数问题,有关这一问题的专业论文往往篇幅冗长,并杂以大量讨厌的公式.我已经省去了证明并对叙述也作了不少简化.最后的几节以简短的篇幅给出了几个绝妙的拼板游戏.虽然这个内容易于理解,对一般的读者也更具有吸引力,但这些内容在马丁的专栏文章与著作中都已有了很完整的论述,所以我无心去超越这位大师.

多连骨牌的发展

多连骨牌位于笛卡儿平面的整点之间,它们由胞腔——即格点

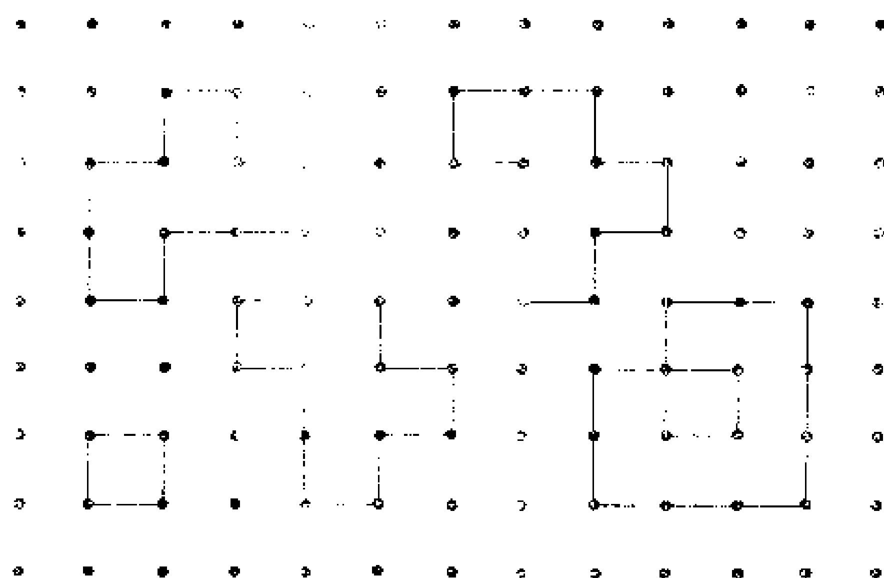


图 1

一个胞腔,F型5连骨牌的两次平移和一次旋转,含有一个洞的7连骨牌.

单位正方形所组成．一个 n 连骨牌是把 n 个胞腔连成一体而构成的，它们的内部互相连通．所罗门·果隆姆把 n 连骨牌描述成国际棋盘上由 n 个方格组成的块，其中任意两个方格之间存在"车"的通路．图1中显示了1个胞腔，3个5连骨牌和1个7连骨牌．图1中所显示的7连骨牌表明，多连骨牌可能并不是简单连接，即中间可以有洞．

图1也说明了另一个问题，就是尽管从外形上看，n 连骨牌的个数有限，但实际上对每个 n 都有无限个 n 连骨牌．千变万化的 n 连骨牌可以按平面上的某些运动群划分为若干个等价类．首先，**平移型**定义为：如果一个骨牌通过平移可以完全覆盖在另一个骨牌上，那么认为这两个骨牌是等价的，即看作是属于同一个平移型．例如，图2显示了4连骨牌的所有不同的平移型．**旋转型**则是这样的：如果一块骨牌通过一系列旋转可以完全覆盖到另一块骨牌上去，那么认为这两块骨牌属于同一个旋转型．由于一次平移过程可以通过两次适当的旋转来完成，所以平移型比旋转型更为精密．图3表明，旋转型骨牌的代表可以从平移型骨牌的代表中挑选出来．从直觉上说，旋转型的 n 连骨牌就是单侧的智力拼板，每一拼块不允许翻转，但可以在一个

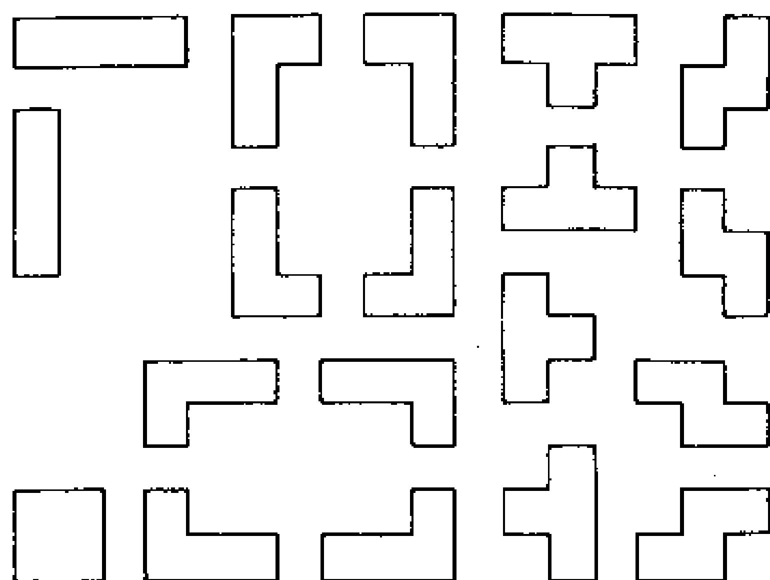


图 2

19种平移型4连骨牌．

平面上自由地转来转去.最后,可以用两个骨牌可否合同,来定义一种等价.由于每一组合同都是由反射合成的,所以两个 n 连骨牌合同指的是可以通过一系列反射,把一个转变为另一个.彼此合同的 n 连骨牌形成反射型.5 种反射型 4 连骨牌显示于图 4 之中.

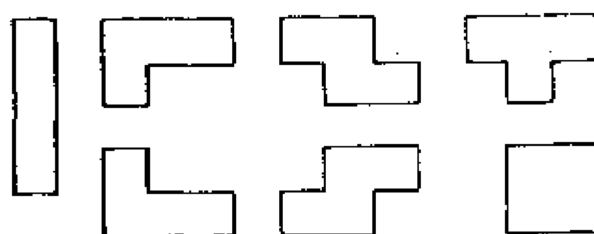


图 3

7 种旋转型的 4 连骨牌.

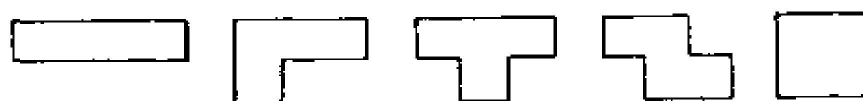


图 4

5 种反射型 4 连骨牌.

计算机被用来统计当 n 较小时(如 $n\leqslant24$)所有的 n 连骨牌的数目,但是 n 连骨牌的数目随着 n 的递增而增大得如此之快,如要罗列全体 n 连骨牌,连我们的太阳系都容纳不了它们,即使是中等大小的 $n(n\geqslant30)$ 也是如此.罗纳德·里凡斯特(Ronald Rivest)建议通过一个巧妙的算法来生成根平移型 n 连骨牌.所谓形成一个根平移型 n 连骨牌,就是在原来骨牌的 n 个胞腔中的一个上点一点("肚脐眼"),以示区别.根平移型 n 连骨牌的数目是平移型 n 连骨牌数的 n 倍.里凡斯特算法的基本点就是定义一次性生成每一个根平移型 n 连骨牌的生成过程.要生成一个根平移型 n 连骨牌,首先是在带点的方格上标上 1,然后从带点的方格上方开始,按顺时针方向逐一在周围方格里标上 2,3,4,5.与多连骨牌方格相邻的方格被称作*边缘方格*.形成根多连骨牌的过程是一个连续的过程,每一步都包含了一个多连骨牌

以及那些标上尽可能小的整数的边缘方格. 任何一个较大的多连骨牌都是通过增加属于这个 n 连骨牌的边缘方格而逐步形成的. 这样，新的未标好的边缘方格总是不断地出现，它们将被标上还没有使用过的最小整数. 这种标号是以带"肚脐眼"的那个方格开始，并按顺时针方向进行. 可以想象，人们必须不止一次地围绕那个带"肚脐眼"的方格，才能把所有的边缘方格标上号码. 图 5 显示了形成一个特定的根 9 连骨牌的全部步骤.

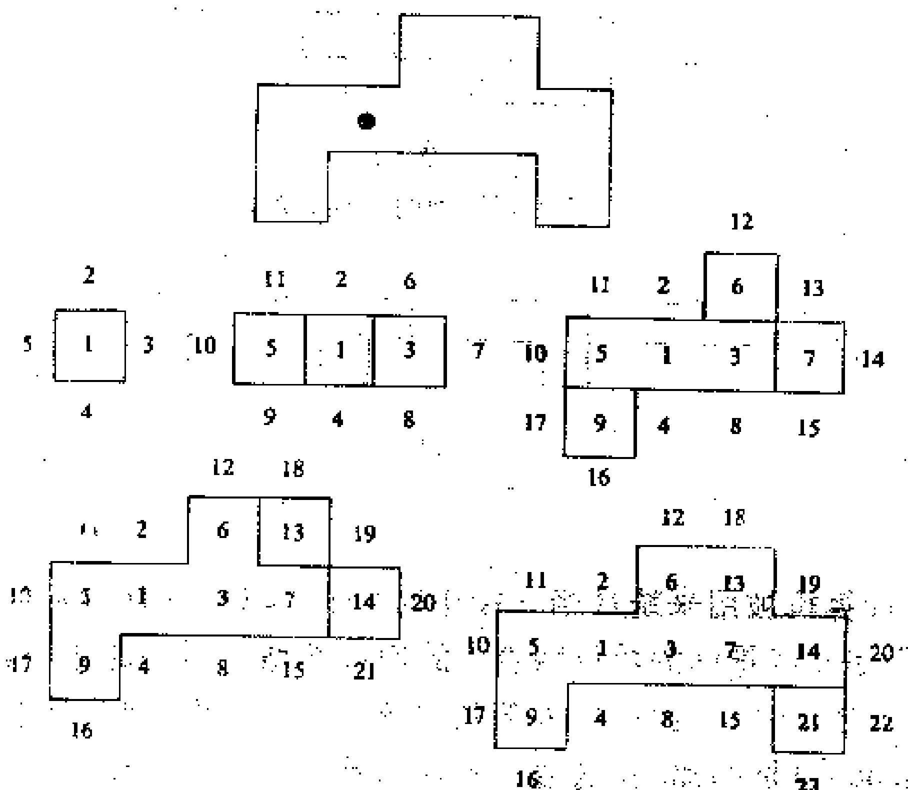


图 5

形成 9 连骨牌的里凡斯特生成算法.

里凡斯特的生长模式揭示了关于根平移型 n 连骨牌的家系树. 凡是从属于某一根多连骨牌 A 的方格的数码集合 $\rho(A)$ 称为这一 n 连骨牌的名称. 例如，图 5 所示的 9 连骨牌的名称为 $\{1,3,5,6,7,9,13,14,21\}$. 若 A 的名称是 B 的名称的一部分，则称 B 是 A 的后裔，即 $\rho(A)\subseteq\rho(B)$. 图 6 显示了根平移型多连骨牌的家系树的一部分.

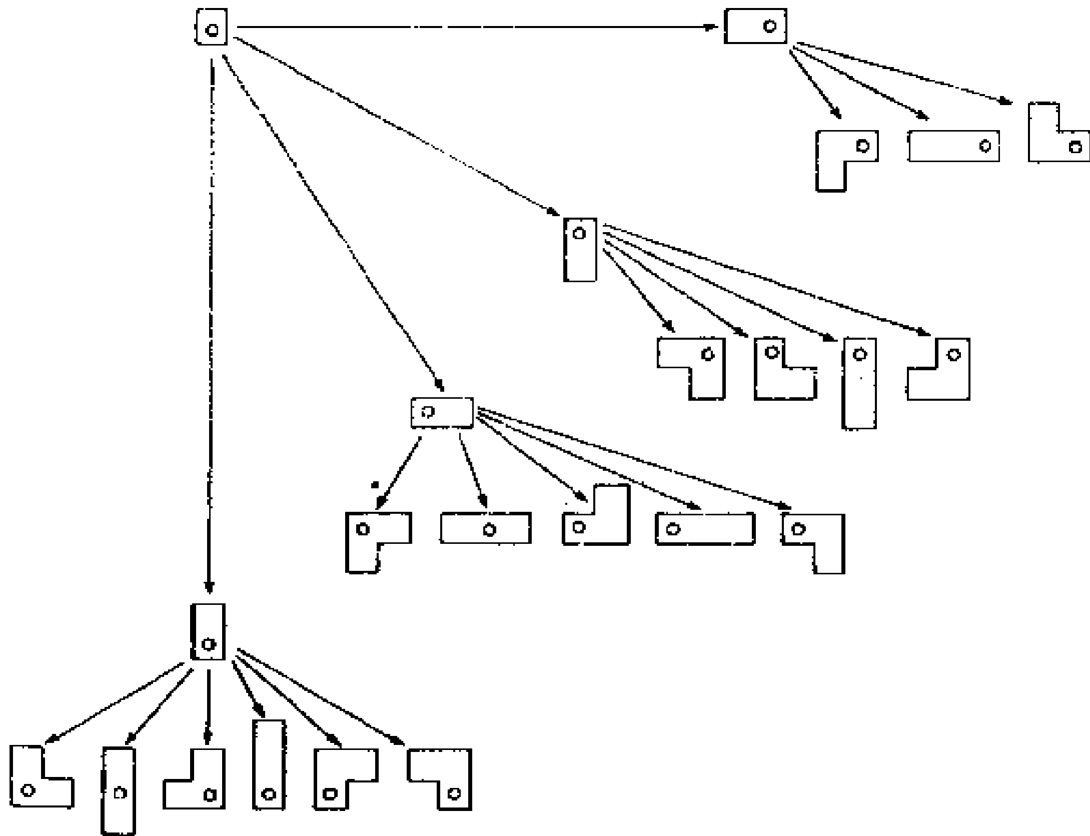


图 6
根平移型多连骨牌的家系树.

如果 n 大于 1，那么 $3n-1$ 是任一根 n 连骨牌的名称中出现的最大整数. 这可用下面的推理来说明：用来标志一个根单连骨牌的边缘方格的最大整数是 5. 如果给原来的根多连骨牌加上一个方格，那么增加的边缘方格为 3（至多）. 因此，用来标志一根 n 连骨牌边缘方格的最大整数最多是 $3n+2$. 如果 $n>1$，那么根 n 连骨牌是在某一根 $n-1$ 连骨牌的基础上加上一个边缘方格而形成的. 那么用来标志这个边缘方格的整数至多为 $3(n-1)+2=3n-1$，因此，由于一个根 n 连骨牌是通过加上标号越来越大的整数方格而形成的，于是所用到的最大的标号为 $3n-1$. 这表明：每个根平移型 n 连骨牌的名称都是从集合 $\{1,2,\cdots,3n-1\}$ 中取 n 个元素所形成的子集. 当然，每个名称

中都必须含 1，所以根平移型 n 连骨牌一一对应于从含 $3n-2$ 个元素的集合 $\{2,3,\cdots,3n-1\}$ 中取出的含 $n-1$ 个元素的子集. 这些子集的数目为 $\binom{3n-2}{n-1}=\frac{(3n-2)!}{(2n-1)!\ (n-1)!}$. 如果 $t(n)$ 表示平移型 n 连骨牌的数目，那么 $nt(n)$ 就是根平移型 n 连骨牌的数目. 上述论证表明：

1. $$t(n)\leqslant\frac{1}{n}\binom{3n-2}{n-1}$$

容易证明

2. $$\frac{1}{(n+1)}\binom{3n+1}{n}\bigg/\frac{1}{n}\binom{3n-2}{n-1}=\frac{3(9n^2-1)}{2(2n^2+3n+1)}<\frac{27}{4},$$

对 $n=1,2,\cdots$ 均成立，所以通过简单的推理可以导出：

3. $$t(n)\leqslant\left(\frac{27}{4}\right)^{n-1}, \text{或} \quad [t(n)]^{\frac{1}{n}}<\frac{27}{4}.$$

获得 $t(n)$ 指数下界的方法将在下一节讲述. 关于 $t(n)$ 的最好下界在我的博士论文中首先给出. 在那里我证明了：

4. 对一切充分大的 n 均有：$(3.72)^n<t(n)$ 或 $3.72<[t(n)]^{\frac{1}{n}}$.

这些上下界说明：平移型 n 连骨牌的数目依指数增长. 如果 $r(n)$ 表示旋转型 n 连骨牌的数目，$c(n)$ 表示合同型 n 连骨牌的数目，那么：

5. $$\frac{t(n)}{8}\leqslant c(n)\leqslant r(n)\leqslant t(n).$$

这样，每种类型的 n 连骨牌都是以同样的指数比率增长的. 这意味着即使有更有效更巧妙的算法来统计所有平移型 n 连骨牌的数目（对较大的 n，如 $n=30$ 而言）都是徒劳的，因为世界上没有足够的空间来存储其输出数据.

多连骨牌的计数

平移型 n 连骨牌的数目大得没法列举，即便当 n 不太大的时候也如此. 但这并没有否定人们计算出这些数目的可能性. 请回顾一下，$t(n)$、$r(n)$、$c(n)$ 分别代表平移型、旋转型、合同型多连骨牌. 尽管

$t(n)$、$r(n)$、$c(n)$随指数增长，然而上面还是有上界，如 8^n，因此，每一个数字还是不会多于 n 位十进制数，有时也会有一些计算按指数增长的序列的好算法．例如，2^n 所代表的十进制数可通过 $2\log_2 n$ 次乘法表示．同样，可以通过阶为 $\log_2 n$ 次的基本运算来计算斐波那契(Fibonacci)序列 1,1,2,3,5,8,13,21,…的第 n 项(斐波那契序列的第 n 项是其前两项的和)．由于斐波那契序列也是随指数增长的，所以第 n 项大约也有 n 位十进制数．究竟有没有计算 $t(n)$的好算法呢？说得更笼统些，有没有计算 $t(n)$的公式呢？

在某种意义上说，还是有一种计算 $t(n)$的公式的．例如里凡斯特增长过程理论就可以用来计算根平移型多连骨牌的数目，从而 $t(n)$就可以通过根平移型骨牌的数目除以 n 而得到．遗憾的是，这种方法需要 $nt(n)$步，而 $nt(n)$是个以指数增长的大数目．据我所知，所有已知的计算 $r(n)$、$t(n)$、$c(n)$的方法都有这个问题，也就是说，算法的计算次数由初等函数 $r(n)$、$t(n)$或 $c(n)$规定了下界，而不是由 n 的某个多项式规定其上界．这也就不难理解，为什么没有好的方法来计算这些数目，而新近发展起来的计算复杂性理论，离我们解决这个问题所需要的水平看来依然差得太远．

令人惊讶的是，即使是 R・C・里德(R. C. Read)所发现的巧妙算法，运行起来的时间仍按指数增长．里德的算法是计算 $t_k(n)$，$t_k(n)$表示落在一条宽为 k 个单位方格的带子之间的平移型 n 连骨牌数目．这种方法可以用来计算 $t(n)$、$r(n)$、$c(n)$，里德通过手算，大约算出了十来个数据．图 7 给出了宽为 2 个单位的带子之间的平移型多连骨牌的数目，其中 $n=1,2,3,4$．这里附有一张关于 $t_2(n)$的表格．

n	1	2	3	4	5	6	7	8	9	10	11	12
$t_2(n)$	2	3	6	11	20	37	68	125	230	423	778	1431

也许，在研究了上面有关 $t_2(n)$的表后，读者会注意到：从第四项起 $t_2(n)$等于前面三项之和，也就是：

1'.　　$t_2(n+3)=t_2(n+2)+t_2(n+1)+t_2(n)$，

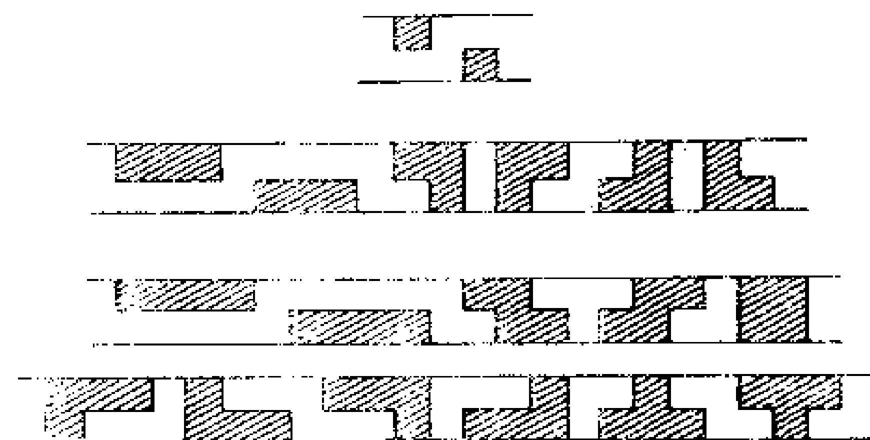


图 7

位于宽为2个单位的带子之间的平移型多连骨牌.

$n=1,2,3,\cdots$. 这个公式说明序列 $t_2(1), t_2(2), t_2(3), \cdots$ 满足一个三阶差分方程. 里德还证明了更一般的结果：对于每个 k，序列 $t_k(1)$，$t_k(2)$，…满足一个差分方程，其阶数依赖于 k. 里德的证明中给出了计算这一差分方程的算法. 他还使用了方阵 M_k 的乘幂形式来表示他的算法. 为了计算 $t_k(n)$，必须计算出 $M_k, M_k^2, \cdots, M_k^{n-k+1}$. 遗憾的是，$M_k$ 的大小是 3^k 阶的，而且必须对不超过 n 一半的全体 k 和 j 都计算出 M_k^j. 这意味着里德的算法具体执行时仍然需要指数时间. 如果用计算机实施里德的算法，能不能算出许多 $t(n)$、$r(n)$、$c(n)$ 的新值来，我对此深表怀疑.

其他平移型 n 连骨牌的子集也被人们考察过. 例如默里·伊登(Murray Eden)，这个问题上的一位先驱者，考虑过子集 $b(n)$，它代表每行都是一条相连的单位方格的那类平移型 n 连骨牌的数目. 这类 n 连骨牌看上去像块不规则木片(图 8)，下面是一张 $b(n)$ 值的表：

n	1	2	3	4	5	6	7	8	9	10
$b(n)$	1	2	6	19	61	196	529	1517	4666	14827

伊登能够证明对一切充分大的 n，有 $(3.14)^n < b(n)$. 当我回到阿

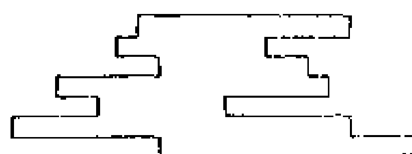


图 8

一块典型的不规则木片

尔伯特大学开始写毕业论文时，李奥·莫塞建议我改进它的限界，我可以证明：

2′. $b(n+4)=5b(n+3)-7b(n+2)-4b(n+1)$,

$n=1,2,\cdots$. 由此可以推出对于一切充分大的 n 有：$(3.20)^n<b(n)$. 大约十年以后，唐纳德·克努特提出不规则木片是"行凸型"n 连骨牌，而凸型 n 连骨牌应当定义为"行"与"列"皆为凸型的不规则木片. 罗纳德·里凡斯特和我发现了一种计算 $d(n)$ 的方法（$d(n)$ 代表凸的平移型 n 连骨牌的数目），并且证明了 $(d(n))^{\frac{1}{n}}$ 趋近于 2.309138…. 埃德·本特(Ed Bender)在我们公式的基础上导出了一个计算 $d(n)$ 的渐近公式.

估计 n 连骨牌数

前面几节表明平移型多连骨牌数 $t(n)$（$r(n)$、$c(n)$ 也一样）随着 n 按指数增长. 更进一步，我们说明了关于计算 $t(n)$ 的已知算法都需要指数时间. 也许这是由于年老与越来越悲观的原因，但我认为没有好的 $t(n)$ 的计算公式是完全有可能的. 实际上，数 $t(n)$ 对于 n 来说（如 $n=30$ 左右）已是"不可知的"了. 这意味着什么？那种认为 $t(30)$ 等于一个通常的十进制数（少于 30 位）的论断必须得到个证明. 或许并没有一个简短的证明？也就是说，所有的证明是那么的长，以致于即使用最大最快的计算机在少于一千年的时间内不停地工作也无法获得——一个多么暗淡的前景！

在面临一个困难问题时，最好从简单一些的问题着手（当然是存

在一定联系的).在我的博士论文中,我证明了当 n 趋向于无穷大时,$(t(n))^{\frac{1}{n}}$ 趋向于一个极限 θ.进一步说,序列 $t(1),[t(2)]^{\frac{1}{2}},[t(3)]^{\frac{1}{3}},\cdots$ 递增地趋于 θ,所以对逐渐增大的 n,计算 $[t(n)]^{\frac{1}{n}}$,可以从下面去逼近 θ.很遗憾,这里需要计算 $t(n)$,而它并没有什么已知的好办法.同时,这个序列收敛于 θ 的速度似乎也很慢,所以为了改进下界 $3.72<\theta$ 将需要一个极大的 n 值.而这一下界我是用一种完全不同的方法建立起来的.

里凡斯特和我设计了一种由上界的递减序列来计算 θ 的方法,可是这个上界的递减序列对计算机来说仍是按指数增长的,并且我们未能证明这个序列真正收敛于 θ.我们所能获得的最好上界为 $\theta<4.65$.最近,帕特·伍德沃思(Pat Woodworth)和我改进了这个办法,看上去似乎可以得到一个的确收敛于 θ 的上界的递减序列.到目前为止,我们还没有用计算机来实现我们的办法.

因此,θ 还是难以捉摸.θ 的小数展开式中连一位都不知道!看来很可能没有什么好的办法来计算 θ,除了前面的一、二位小数之外,θ 的其他小数也许是不可知的(谈论这个问题时,我有时开玩笑地把 θ 称为克拉纳常数.没有一位数是知道的,并且也许永远是个谜!如果可以证明 $\theta=4$ 那么是否可以让小学生数:"一、二、三、克拉纳常数、五、……?"哈哈).

以下将用已知的 $c(n)$(反射型多连骨牌的数目)数值的表格来结束本节.这些数字(以及其他类型的多连骨牌数)是由 D.H.莱德梅耶(D.H. Redelmeier)计算而得的.有一个关于这个序列的著名猜想:比值 $c(2)/c(1),c(3)/c(2),\cdots$ 是递增的,即

$$\frac{c(n)}{c(n-1)}<\frac{c(n+1)}{c(n)},\qquad n=2,3,\cdots.$$

如果这个推测是正确的,那么这个序列将递增地趋向于克拉纳常数 θ.进而言之,每一个这样的比率总是 θ 的一个下界,特别地,$\frac{c(24)}{c(23)}<\theta$,也就是说,$3.8977\cdots<\theta$,那就比现有的下界 $3.72\cdots<\theta$ 要好,当然这个改进是要取决于上述猜想是否真正成立.

n	$c(n)$	n	$c(n)$	n	$c(n)$	n	$c(n)$
1	1	7	108	13	238591	19	742624232
2	1	8	369	14	901971	20	2870671950
3	2	9	1285	15	3426576	21	11123060678
4	5	10	4655	16	13079255	22	43191857688
5	12	11	17073	17	50107909	23	168047007728
6	35	12	63600	18	192622052	24	654999700403

表 1

不同的 n 连骨牌数.

多连骨牌的拼板游戏

马丁·加德纳 1958 年发表的关于多连骨牌的文章把我引向了这个课题，那时，我还是个高中学生. 这篇文章的特色在于所罗门·果隆姆所发明的一些游戏及其证明. 文章特别指出，十二块反射型五连骨牌（下文将简称五连骨牌）可以组成面积为 60 的各种不同矩形，其形状大小为(3×20)、(4×15)，(5×12)和(6×10). 几年以后，C·J·博坎普(Chris. J. Bouwkamp)运用计算机列出了用十二块多连骨牌解决装箱问题的所有方案. 除去这些已经列出的两维装箱问题，博坎普还装成了形状大小为(2×3×10)，(2×5×6)和(3×4×5)的箱子. 可是在 1958 年，当时马丁的读者们却不可能从博坎普所列的方案或者他的巧妙的计算机程序中捞到好处，事实上，我也是花了很长一段时间艰难地寻找用五连骨牌组成(3×20)长方形的两种方法，并且是把解法送给马丁的许多热心读者当中的一个. 这两种方法（都是唯一的）列于图 9 之中.

在我准备这篇文章的时候，我意识到还没有人考虑过旋转型多连骨牌的问题，也就是，列于图 10 之中的 18 个"单侧"五连骨牌. 但是，果隆姆在他们的《多连骨牌》一书中通过讨论这些骨牌如何组成

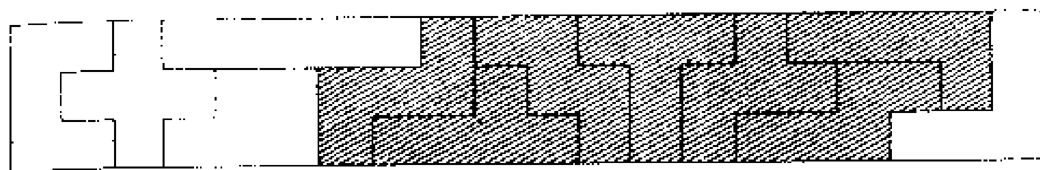


图 9

用五连骨牌装满 3×20 矩形的两种办法，图中阴影区域绕中心旋转 180°后即得第二种解法.

(9×10)矩形已涉及这个问题. 也许本节所讨论的另外一些问题果隆姆也涉及了. 如果读者打算做一套 18 块的单侧多连骨牌，那最好用纸板，一面着白色，另一面着黑色，当拼板的时候，必须把相同颜色，例如黑色，放在一面.

由于一共有 18 块旋转型五连骨牌，它们一定能组成总面积为 90 的矩形. 有 4 种大小的可能性，就是(3×30)，(5×18)，(6×15)和(9×10). 按从"瘦"到"胖"的顺序，各个矩形的排法会变得愈来愈容易. 比如，我拼(3×30)矩形花了 2 小时，而拼(9×10)矩形只花了 2 分钟. 这种排法难易的顺序也反映出一个事实，那就是(3×30)矩形的排法只有几种，而(9×10)矩形的排法有几千种. 毫无疑问，克利思·博坎普将成为用定量办法确认这个观点的第一人. 图 11 显示了问题的解决方法.

看上去用单侧五连骨牌拼成(3×30)矩形确实很难. 但是，一旦你已经发现了一种排法，那么通过旋转或内部交换往往可以把一种方法转变为另一种新方法. 为了图示这种旋转方式，请看图11中由6至13骨牌拼成的那一大块，如果把这一大块以其中心旋转180°，就可以得到图12中的第一种解法. 再如，把图11中的6至11及15至17所拼成的两大块对换一下，就形成了图12中的第二种方法. 这里一共列举五种由图11变化而得到的不同类型的拼法，其中两种是旋转，其余三种是互相对换. 当然这种变化方式同样可以运用于图 12 之中，结果

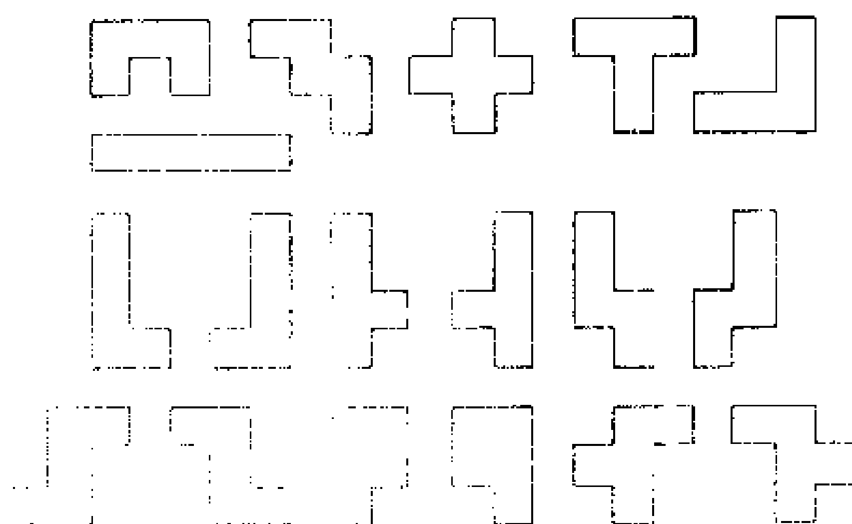


图 10

旋转及反射下不等价的 18 种类型.

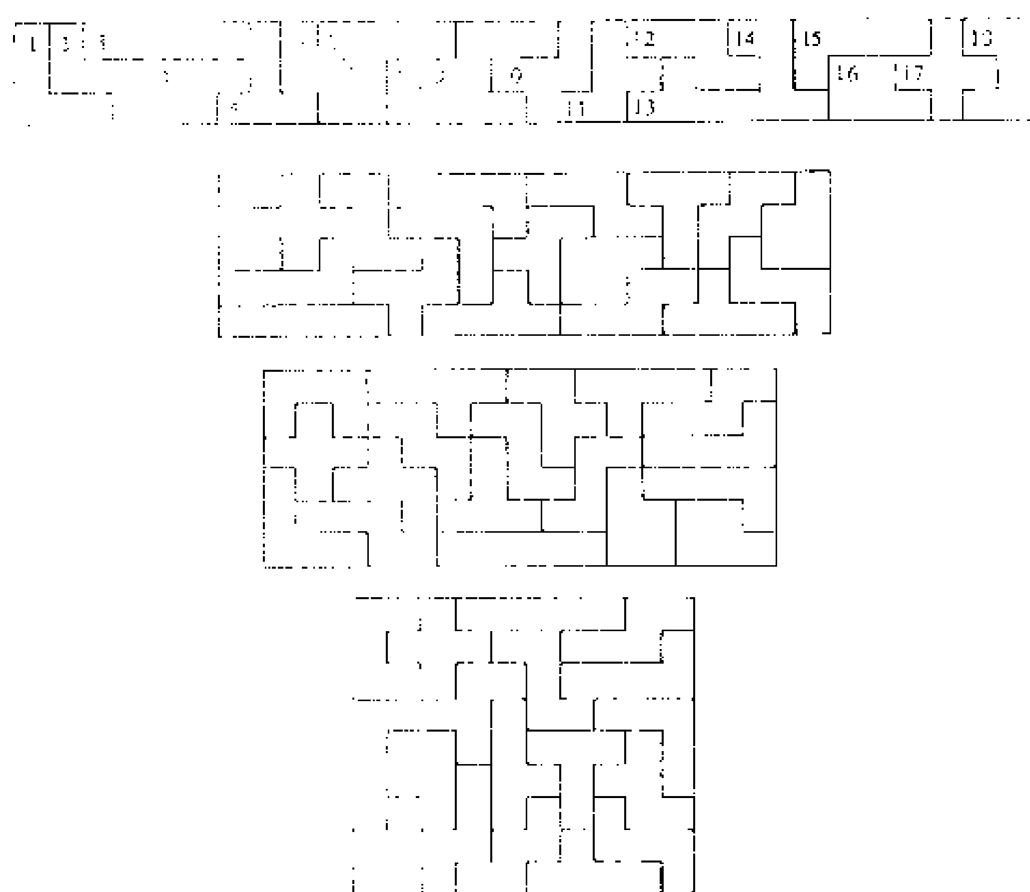


图 11

用单侧五连骨牌拼成的各种矩形.

能得到 9 种新的拼法. 或许读者有兴趣知道通过旋转和对换究竟能得到多少种排法. 当然, 任何一种排法都可以通过反射而得出新的拼法. 然而这种反射办法所得出的拼法是不应当看做不同的.

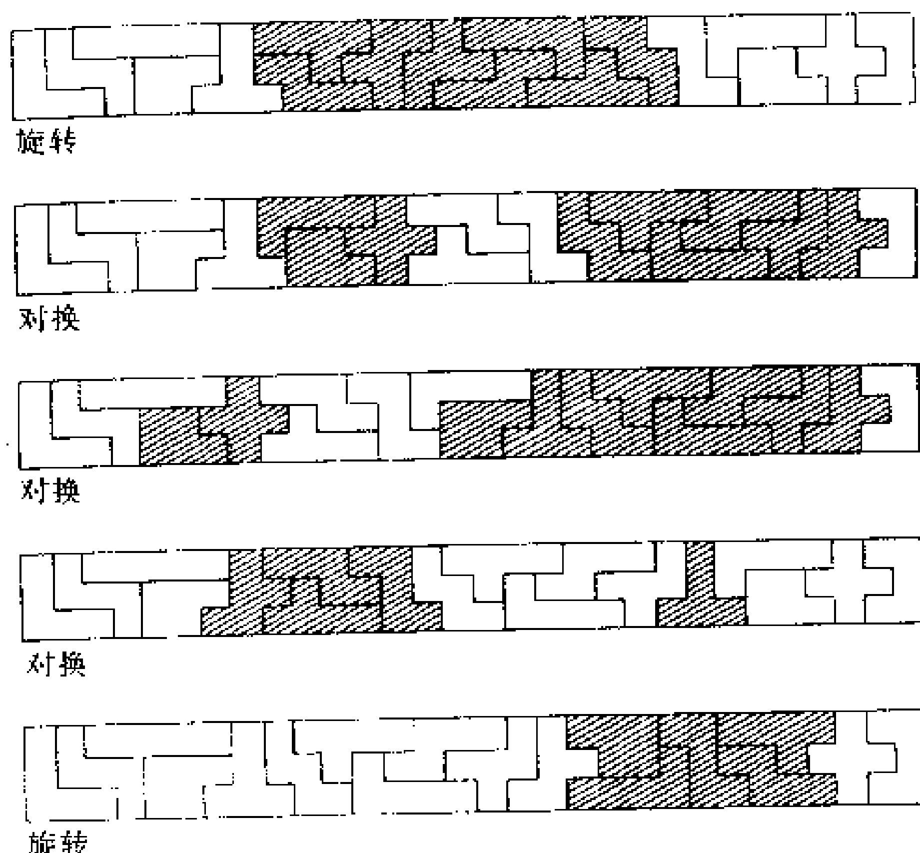


图 12

用图 9 中得到的单侧五连骨牌拼成 3×30 矩形.

在我研究多连骨牌的早期, 我着力于找定义得巧妙的拼块集合, 其后又把注意力转变到用这些拼块还能进一步干些什么上去了. 这类事已在上文通过单侧五连骨牌来说明过. 拼块集合的奥妙在某种意义上来说, 在于其定义越简单越好. 这些年来, 我一直在问是否有更简单的多连骨牌也能产生复杂问题. 我自己的发明是运用一块特殊的多连骨牌或多连立方体的几份拷贝(多连立方体是多连骨牌在三维空间的推广). 现在的问题已经成了怎样用一块骨牌的拷贝来拼矩形. 下一节将给出它的一个实例.

“N”型五连立方体

图 13 中显示了一个 N 型五连立方体——它的组成形状说明了名字的由来. 很显然，N 型五连立方体就是由 N 形的 5 连骨牌的五个单位方格用五个单位立方体替换而成的. 很容易证明 N 型五连骨牌不能拼成一个矩形，因为用 N 型五连骨牌不能组成长方形的一条边. 因此，用 N 型五连立方体也不能拼成其中一边为一个单位的盒子. 那么用 N 型五连立方体究竟能否拼成一个盒子呢? 回答是肯定的. 我将描述如何拼成有一条边为 2 个单位的盒子.

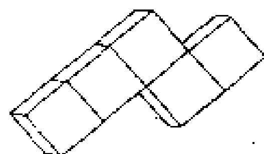


图　13

N 型五连立方体.

如果一个 $a\times b\times c$ 盒子可以由五连立方体装成，那么每个五连立方体体积(即 5)应能整除整个盒子的体积 abc. 这意味着整数 a、b、c 中的一个是 5 的倍数. 考虑当 $a=2$ 时的特例，那么 b、c 中至少有一个应是 5 的倍数(比如 c). 很容易证明，当 $b=1$、2、3 时，即使 c 为 5 的倍数，N 型五连立方体也不能拼成 $2\times b\times c$ 的盒子，但是拼成$(2\times4\times5)$，$(2\times5\times5)$$(2\times6\times5)$、$(2\times7\times5)$的盒子是可能的. 如图 14 所示. 这些盒子又可能用来拼成较大的盒子，很容易证明，每个大于 3 的整数 b 都可以用表达式 $b=4w+5x+6y+7z$ 表示，其中 x,y,z,w 为非负整数. 所以包含一面为 2×5 的盒子 $w(2\times4\times5)$，$x(2\times5\times5)$，$y(2\times6\times5)$，$z(2\times7\times5)$可以集在一起组成 $2\times b\times5$ 类盒子，由 N 型五连立方体拼成. 在这些盒子中，我们可以根据 $2\times b$ 面找一个匹配

数 k，拼成 $2\times b\times 5k$ 盒子．结论是，包含能够用 N 型五连立方体拼成的 $2\times b\times c$ 盒子的集合称为 S_2，其中 b 必须大于 3 而 c 是 5 的倍数．进一步说，图 14 中所显示的 4 个立体就是 S_2 的一个基，也就是，每一个能用 N 型五连立方体拼成的 $2\times b\times c$ 盒子一定能用图 14 所显示的四个立体拼成．

集合 S_2 包含了所有可以用 N 型五连立方体拼成的二层型盒子．这些盒子可以进一步拼成具有偶数层的（大）盒子．这意味着图 14 所显示的 4 个立体也可以是当 $a,c=1,2,3,\cdots$ 以及 $b=4,5,6,\cdots$ 时 $2a\times b\times 5c$ 立体的一个基．这类盒子不可能是 3 层，那么能否用 N 型五连立方体拼成 3 层的盒子呢？到目前为止，我已经发现的拼法有：$3\times 5\times 8$ 盒子（见图 15），$3\times 4\times 5k$ 盒子，k 取 2，3，…（见图 16）．当 k 大于 1 时，所有的 $3\times 4\times 5k$ 盒子可以由 $3\times 4\times 10$ 和 $3\times 4\times 15$ 盒子拼成．已经证明，$3\times 4\times 5$ 盒子不能由 N 型五连立方体拼成．现在还不知道哪一类 $3\times 5\times n$ 盒子可由 N 型五连立方体拼成，但是有一个算法可以解决这个问题．

每一个可以由 N 型五连立方体拼成的 4 层盒子均可由已形成的 2 层或 3 层的盒子拼成．那么 5 层盒子如何呢？关键问题是能否拼成边为 5 的正方体．我在仔细考虑以后发现了这种拼法，见图 17．实际上，当最后一块骨牌归位时，我已经精疲力竭，几乎晕过去．

这里列举了足够多的可由 N 型五连立方体拼成的盒子，说明当 a、b、c 足够大，并且 abc 为 5 的倍数时，任何一个 $a\times b\times c$ 盒子都能拼出来，正如下一节中的讨论所表明的，我们预期会有这些结论．那些可以拼起来但不能分割得更小的盒子称为既约的拼块（犹似自然数中的素数）．既约块的集合是有限的，但对 N 型五连立方体来说，情况不明．

用“砖块”拼成盒子

在我居住在荷兰与德·布鲁因(de Bruijn)一起工作的日子里，

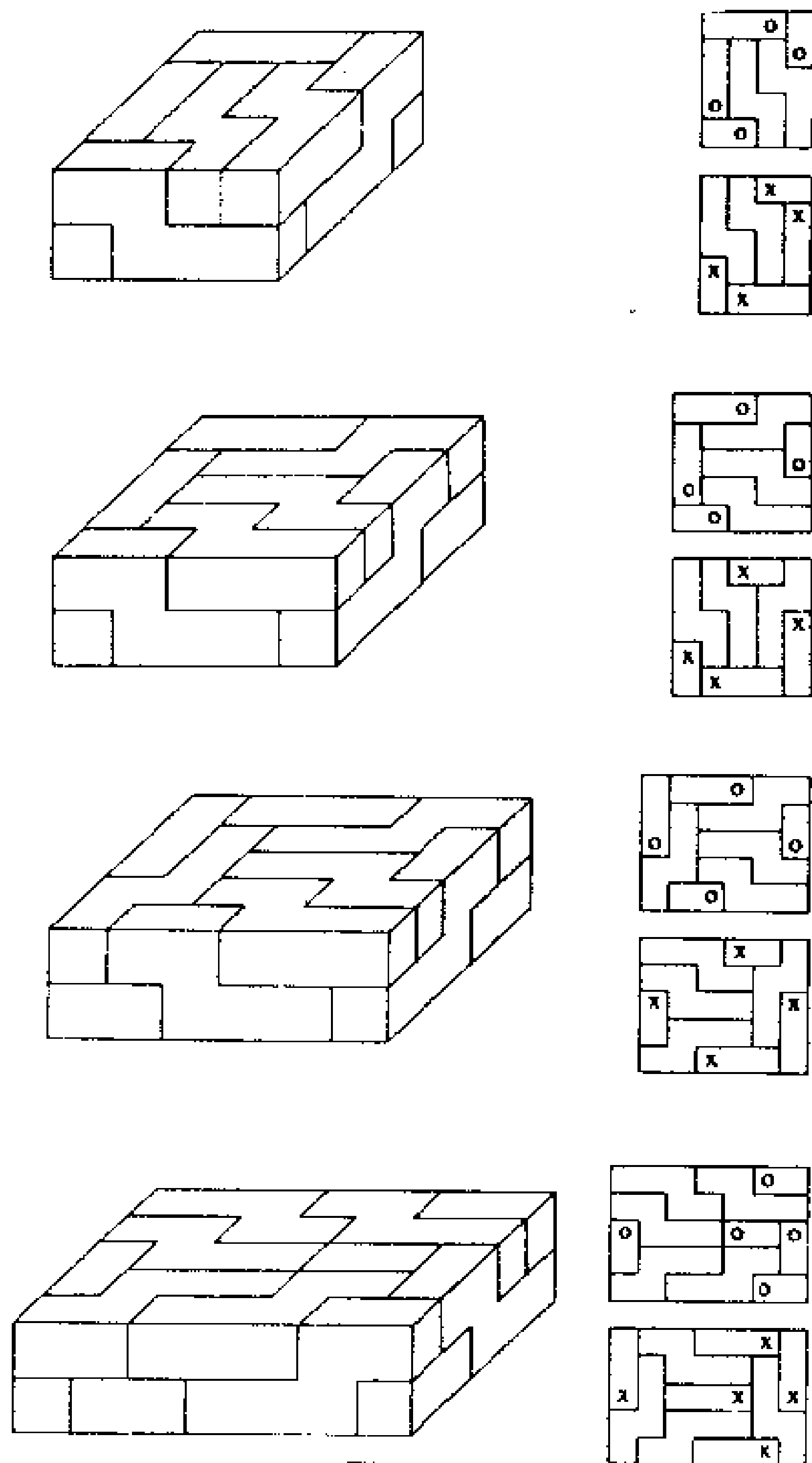


图 14

四个N型五连立方体所拼成的既约块．上层中打圆点的单元应与下层中打×的单元配合．

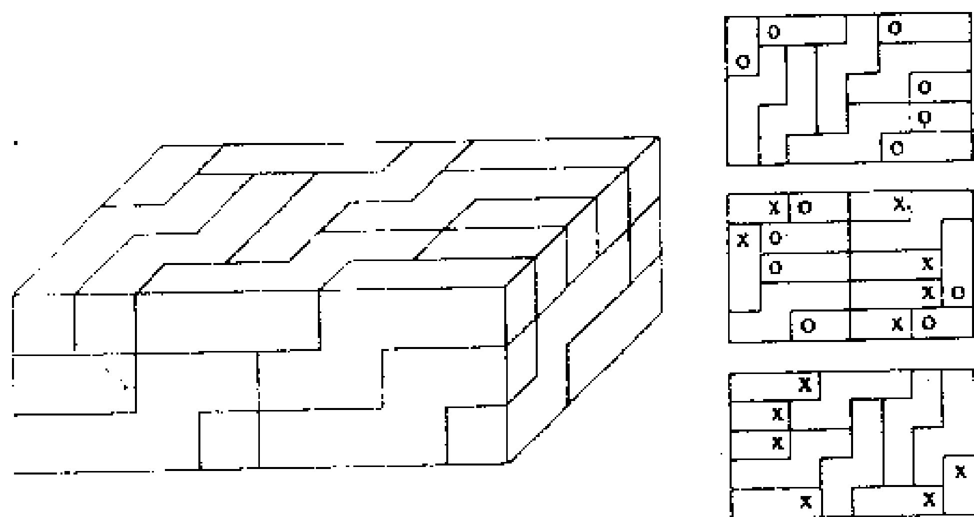
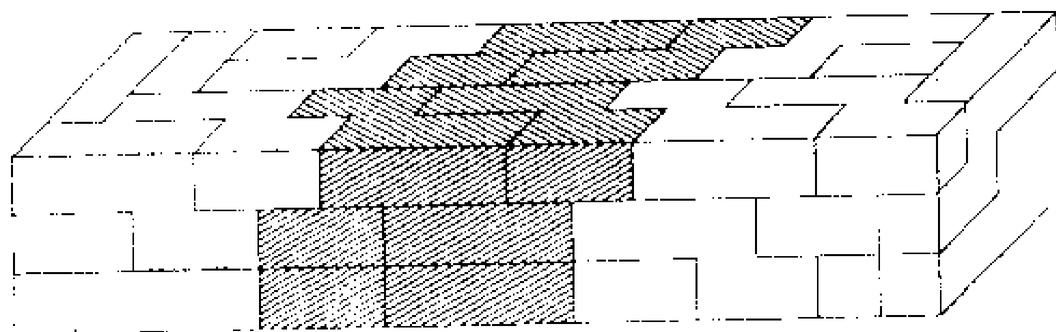


图 15

用N型五连立方体拼成的3×5×8盒子.



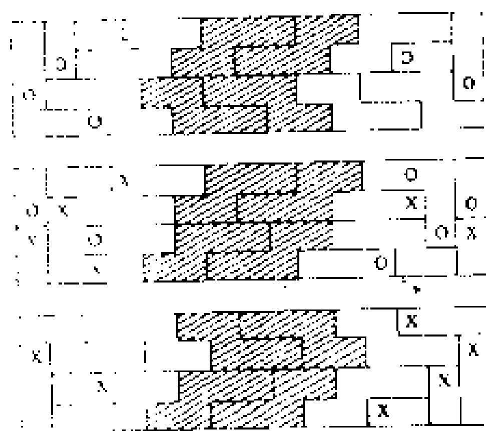


图 16

用N型五连立方体拼成的3×4×15盒子，如拿掉图中的阴影部分，把其余部分推挤一起，便可得到一只3×4×10盒子.

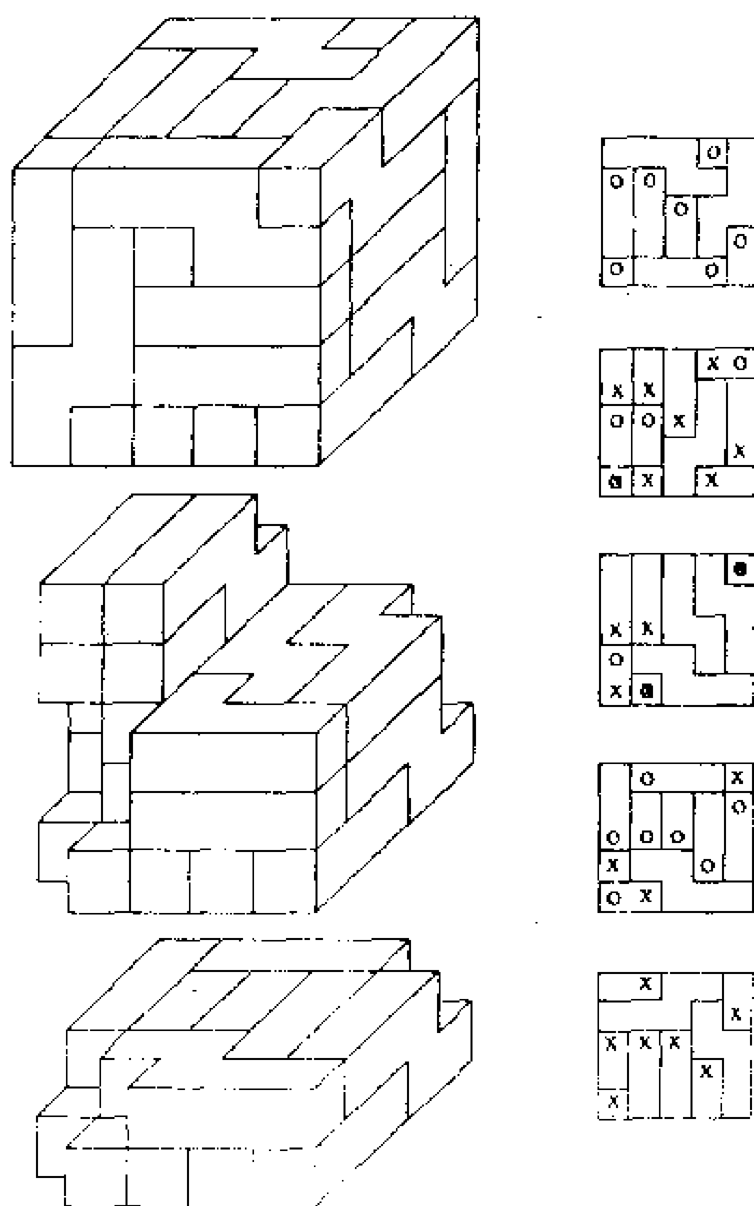


图 17

用N型五连立方体拼成的、棱长为5的立方体.

我有许多收获. 其中之一就是认识了弗里茨·甘倍尔(Frits Göbel). 他是另一位多连骨牌的研究者. 那时候, 弗里茨有一本笔记, 收集了一些有关这个课题的信息、问题和猜测. 他让我回想起果隆姆提出的而我在《美国数学月刊》上解答过的问题, 用简括的话归纳这个结论就是: 每一个可以用L型四连骨牌拼成的矩形一定能由2×4和3×8矩形拼成, 而2×4和3×8矩形本身可以由L型四连骨牌拼成. 这

个问题的另一个结论由戴维·瓦尔科普(David Walkup)得到，即每一个可以由T型四连骨牌拼成的矩形一定可以由边长为4的正方形拼成. 当然，边长为4的正方形也可以由T型四连骨牌拼成.（由L型四连骨牌拼成的2×4和3×8矩形和另一个由T型四连骨牌拼成的4×4正方形，见图18.）鉴于此类发现，弗里茨推测可以由给定多连骨牌拼成的矩形集合 R 有一个有限的基 B，也就是，R 有着有限子集 B，并且 R 中的每个矩形可以由 B 中的矩形拼成. 弗里茨还推测，如果矩形 A 与 B 的面积不具有公共素因子，则任何边长为充分大的矩形都可由 A 和 B 拼成. 他想用这个结论证明前面的推测. 但后来发现第二个推测是错误的.

受甘倍尔推测的启发，我发现了 $a\times b$ 矩形可由 $p\times p$ 和 $q\times q$ 正方形拼成的充分必要条件. 这类矩形或者可以直接由一种大小的正方形组成，或者它能分解成两个较小的矩形，每个小矩形可由一样大小的正方形组成. 这意味着下列几项中必有一项是正确的：1)a、b 都能被 p 整除. 2)a、b 都能被 q 整除. 3)a 或 b（比如 a）可以同时被 p 和 q 整除. 此时，$b=px+qy$，x、y 是非负整数. 这说明，如果 $p=2$，$q=3$，那么正方形 5×5，25×25，125×125，…，$5^k\times5^k$，… 中没有一个能由

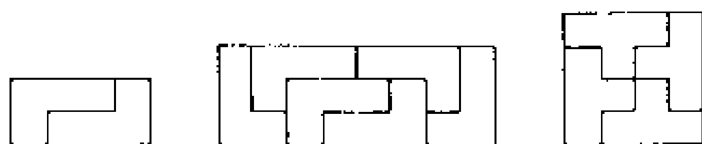


图 18

由 L 型四连骨牌与 T 型四连骨牌所拼成的"既约块".

2×2 正方形或 3×3 正方形拼成. 当然，这就推翻了甘倍尔的第二个推测，也就是有关第一个推测的一种证明手段. 然而第一个推测是正确的. 事实上，我还能证明一个更要强得多的结果.

假如 R 是由整数边组成的矩形的一个无限集合，这些矩形的特点是它们固定在平面上并具有定向性. 也就是，一个 $a\times b$ 矩形长的

一边平行于 y 轴而另一边平行于 x 轴. 因此，如果 $a \neq b$，那么 $a \times b$ 矩形与 $b \times a$ 矩形是不同的. 我们认为通过平移重新确定的矩形与原来的矩形是等价的. 另外这种拼法还有一个特点：一个矩形必须先分成两个较小的矩形，然后再用同样的方法拼成大的. 例如，我对由两种正方形拼成的矩形的理论主张：这类矩形可以通过“先分后合”的方法拼成. 但是，用“先分后合”拼法不能拼成图 19 所示的图形.

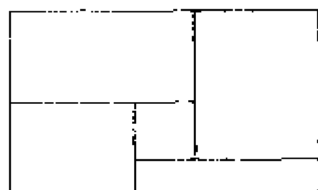


图 19

一种不能用“先分后合”法完成的图形.

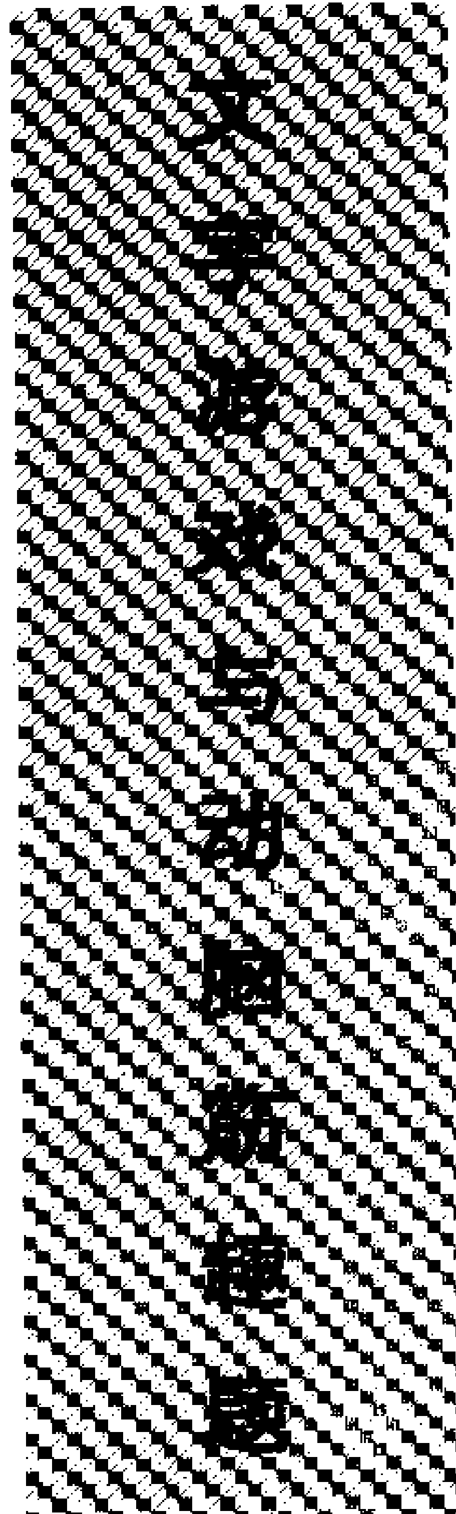
我的理论涉及到用“先分后合”方法去拼定向性矩形. 它认为，每个定向性矩形的无限集合 R 含有一个有限子集 B，而 R 中的每一个元素都可以通过“先分后合”法由 B 中的元素拼成. 进一步说，类似的结论在 k 维空间之中也成立.

狄克・德・布鲁因(Dick de Bruijn)发现了一种用调和块拼盒子游戏的很巧妙的定理. 他把一个调和块定义为一个大小为 $a \times ab \times abc$ 的立体，其中 a、b、c 为整数(对多维立体也有类似的定义). 结论是当且仅当一只盒子可以由所有平行块拼成时它才能由这些调和块拼成，也就是，这些盒子的大小为 $ap \times abq \times abcr$. 这说明当且仅当盒子能由“先分后合”方法拼成时，它才能由调和块拼成. 我已经证明的另一定理断言，当且仅当矩形可以由“先分后合”方法拼成时，它才能由一个矩形的几份拷贝拼成. 这使我进一步推测：在 k 维空间里，只有可以由“先分后合”方法拼成的盒子才能用调和块拼成. 然而戴维・辛马斯特(David Singmaster)提供了一个反面例子：用 $1 \times 3 \times 4$ 拼块可以拼成一个 $5 \times 5 \times 12$ 盒子，但是这个盒子却不能用“先分后

合"法拼成!

这个结论与我对甘倍尔第一个推测的证明相符,也就是:除去盒子的一个有限集合外,其他的盒子都可以用"先分后合"法由一个给定的砖块来拼成.

祝马丁生日愉快!


大学游戏与动脑筋趣题

失踪之谜

● 斯坦福大学

□ 唐纳德·E·克努特(Donald E. Knuth)

请读者注意：把第一首诗右半部的上三行与下四行交换了一下位置后，它变成了只有七行的第二首诗了．请问，失踪的是哪一行？

	我惊讶魔术家们用何妙法使其兔子消失无踪；
变戏法的咒语能够	不抵触
科学上的定律	与明白无误的数学事实.
魔术家们采用狡猾手段	
(但他们永不会告你以真相)	搬运东西如在梦里；
时而	挂起来，赶出去，或干脆无效验-看来是如此.
戏法中必定有什么奥秘	埋伏着
用巧妙手法表演时	一种揭破疑谜的力量.

变戏法的咒语能够	搬运东西如在梦里；
科学上的定律	挂起来，赶出去，或干脆无效验-看来是如此.
魔术家们采用狡猾手段	埋伏着
(但他们永不会告你以真相)	一种揭破疑谜的力量
时而	我惊讶魔术家们用何妙法使其兔子消失无踪；
戏法中必定有什么奥秘	不抵触
用巧妙的手法表演时	明白无误的数学事实.

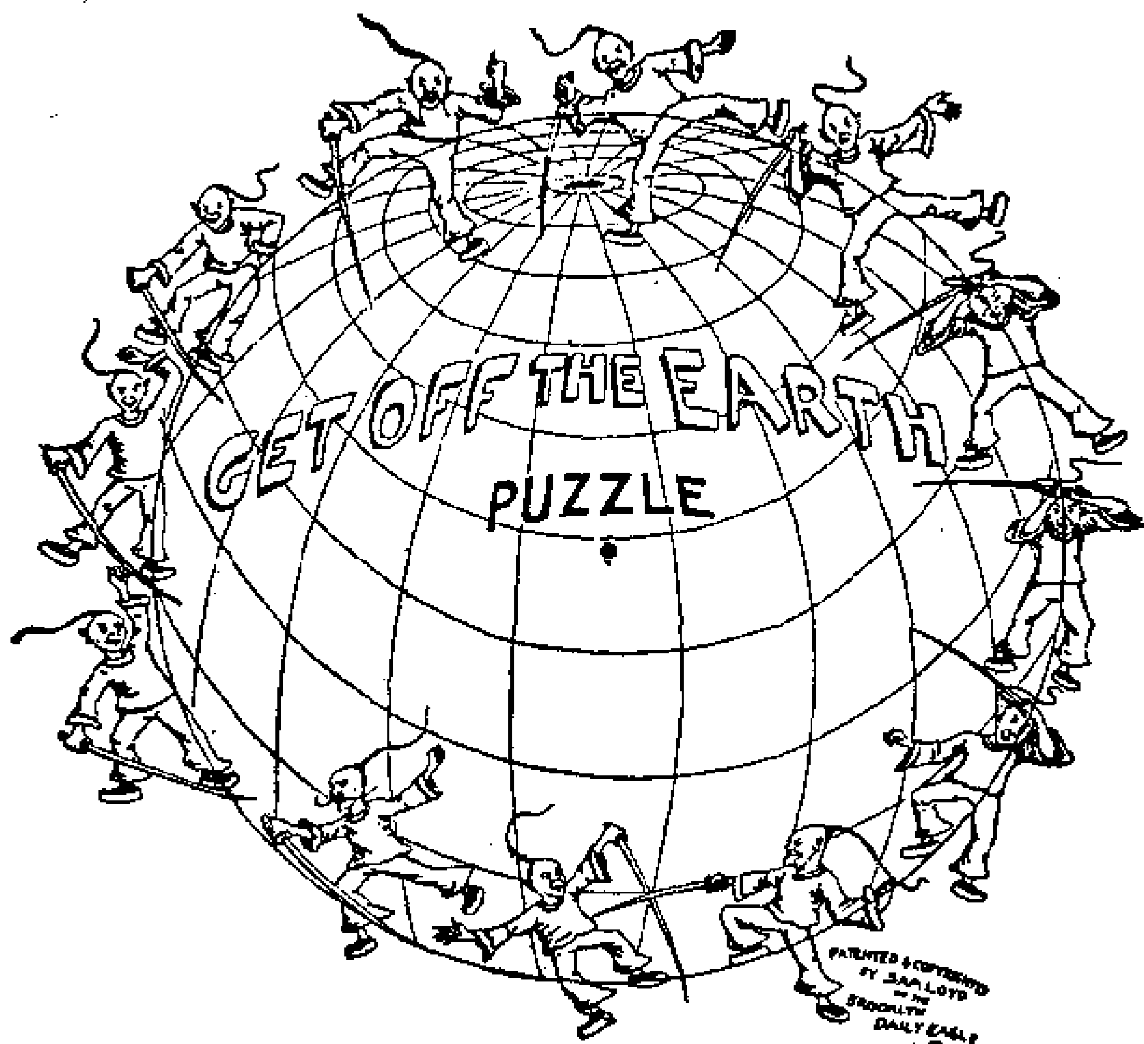
I wonder how magicians make their rabbits disappear;
Enchanted words like "hocus pocus" can not interfere
with laws of science and facts of mathematics that are clear.
The prestidigitators, making use of devious schemes,
(although they never tell you how) transport things as in dreams:
At times suspended, banished, null and void — or so it seems.
There must be something secret, yes, a trick that will involve
— when done with sleight of hand — a force that's able to *dissolve*.

N.B.: When the right-hand portions of this eight-line poem are interchanged,
a seven-line poem results. Which line disappears? —D. E. Knuth

Enchanted words like "hocus pocus" can transport things as in dreams:
with laws of science suspended, banished, null and void — or so it seems.
The prestidigitators, making use of devious schemes, involve
(although they never tell you how) a force that's able to *dissolve*.
At times I wonder how magicians make their rabbits disappear;
There must be something secret, yes, a trick that will not interfere
— when done with sleight of hand — and facts of mathematics that are clear.

译者注：

人们常说“无故失踪”，但事实上，任何失踪都是有“故”的，这种缘故，可能埋藏得很深、很深。设法把它挖掘出来，是许多人类活动之目的，例如犯罪的侦缉，失物的寻找……。有趣的是，数学游戏里头也有这类题材，而且一度曾在美国广泛流行过，这就是计算机科学大师唐纳德·E·克努特(Donald E. Knuth)先生撰写此文的社会文化背景。在马丁·加德纳的一本名著《数学·魔术·奥秘》(Mathematics, Magic and Mystery，因这三个英语单词的第一个字母都是M，故又简称“三M”)中，曾引证过许多材料。据他考证，在这个问题上



附 图

的始作俑者，是十九世纪美国游戏数学大师山姆·洛伊德，他曾为纽约布鲁克林《每日之鹰报》制作过一种玩具，名为“开除球籍”。此物是一个圆形转盘，可以旋转自如，上面画着十三个清朝打扮的人物，手持尖刀，似在决斗、又像练功。但是转到某一位置时，画面上的人物却只有十二个了，那么，还有一人到哪里去了呢？据历史记载，此物曾售掉数百万份之

多，风靡一时（见附图），影响远播日本与东亚.

由于中、英文是两个根本不同的语种，所以无法完全照搬英文中的句子.这里只能做到尽可能最大限度地保留原文的意义与形式.实际上，每种文字都有其特定的词汇、语法与前后搭配关系.将这种文字游戏转变为异国文字，确实是非常困难的，有时甚至是根本办不到的.下面我们略举一例.譬如说，“今年真好晦气，全无财帛进门”是一句很不吉利的对联；但是经过祝枝山重新标点之后，却成了“今年真好，晦气全无，财帛进门”的好口彩了.试问，你能把它全盘变成英语或任何其他文字吗？

因此，为了保持原来的神韵，我们只好把作者的英语原作全部复印出来，以供通晓英语的读者欣赏，好在它并不很长.

非欧和声法[1]

● 斯坦福大学

□ 斯科特·金(Scott Kim)

欧 几 里 得

古希腊人的一桩突出成就是在日常生活经验的主要声音中构筑起一个音调和声的演绎系统.这一系统在欧几里得的十三卷《和声基础》[2](又称"调和级数")中得到了充分阐述.

一上来是五条公理：

1. 若给出两个音调，则必定存在着一个连接它们的音程.(琶音公理)

2. 一个音程可以无限延长.(延长音公理)

3. 基调与一个和音一经定出，即可据以作出进行曲.(顺序公理)

4. 所有的主音调都相等.("魔鬼"的公理)

---

[1] 译者注：本文是一篇诙谐文章，作者堆砌了一些根本无关的音乐术语，以此作为借喻，用来说明非欧几何、四色猜想以及其他问题.

[2] 译者注：实际上欧几里得根本没有写过这本书，所指的乃是他的《几何原本》.作者的此种手法，除个别地方之外，以下将不再一一注明，请读者注意识别.

5. 如一旋律线与伴随它的两条旋律线之间存在着五分之一的差异而使同侧的内部音程小于两个三音调，则在无限延伸时，两旋律线将最终转为音程小于两个三音调的那一侧.（第五公设）

欧几里得能推导出诸如以下的定理：

一个三和弦的各音程之和等于两个三音调.

他也研究过阿波罗尼斯（Apollonius）进行曲，得到以下结论：第五个圆内的任一点都有一个反演对应点.

虽然如此，欧几里得显然不满意他对第五公设所采取的说法.它没有前面四个公理的紧凑与优雅.实际上，欧几里得在其《基础》一书中尽量推迟使用第五公设，而前面四个公设却是开足马力，大用特用.他一度考虑要增补一个公式，即所谓"附加的第五公设"，但最后还是放弃了这一想法，因为它仅不过是一个"小六子"（小小的第六公设）.

其后又经许多世纪，不少人试图挑出欧几里得的每个小毛病.也许第五公设是多余的，有可能利用第五公设作出的一切曲子都可仅由前四个公设即能作出.有人想通过靠不住的节奏（所谓"分解性定理证明法"）来证明这一点.

17 世纪意大利音乐理论家吉洛拉莫·萨凯里（Gerolamo Saccheri）[1]创作了与第五公设针锋相对的一整套乐曲.他希望能找到一种交叉关系，结果什么也没有出现.萨凯里的研究工作是非欧作曲法的几个早期实例之一.然而，他所受的教育却未能使他意识到其发现的重要性.由于未能接受新的高昂音调，他于是宣称，他的作曲法"有悖于乐音旋律的本质"，从而否定了他在历史上应拥有一席位置.就这样，第五公设的解决仍然需要搁置多年.

---

[1] 译者注：此人根本不是音乐理论家而是一位几何学家，请参看《什么是非欧几何？》一书的有关章节.

希腊人的理想

像他的前人一样，萨凯里的失败是由于希腊思想对音乐理论的支配性影响．希腊人抱有一种想法：神圣的比统治着宇宙．同一种比支配着行星的轨道、几何学的和谐乃至于音乐的图式．发现了不少具有简单比例关系的例子．毕达哥拉斯定理说明了斜边所代表的音高与边长之间的关系．如果长度加倍，则音高折半．毕达哥拉斯音调系统奠基于比 3/2，它被称作*黄金音程*，人们认为它听起来最悦耳．

希腊人最终把各种比同泛音序列挂上了钩，这是一个许多世纪以来引起深切反响的课题，人们经常重复这句话："上帝创造了泛音序列，其他东西都是人的产物．"他们从泛音序列推导出不同音调系统，也就是调式．调式逻辑家们相信音调对人的气质具有强烈作用．比与实在不能分离，因而音乐就是宇宙现存秩序的启示．

既然只有一种音乐，那就没有必要把它弄得很清楚．对一位现代音乐家来说，欧几里得的常用记谱法，例如"整体大于一部分"以及诸如此类的东西，听上去很刺耳．他们间接提到其他的记谱法，它们也同样晦涩难懂，因而无法说成是自明的定义．类似地，当追随欧几里得的作曲家们认识到正确领唱的重要性时，他们便感到有必要把可能会引起混淆的步子标记出来．

这种缩略倾向直接导致了一系列引起混乱的记谱练习，其中包括数学虚构，低音部与装饰音．演奏者必须以合适的方式填补空缺的几步．大数学家 J・S・巴赫(J. S. Bach)[1]被认为是用老框框写出所有的装饰音．最微妙的省略体现在器乐演奏的谱曲中．作曲家们时常仰赖于其乐器的特性．如果一位演奏家完全指望通过"乱弹琴"来重建乐音，那就需要更多的信息．

希腊人的理论中也不是没有唱反调的人．批评家们提出了需要

[1] 译者注：巴赫是著名古典音乐家，作者存心把他说成是数学家．

作出解释的若干基本问题，其中有一些只是在现代实践中才能判定是非. Alea 地方的 Zeno 提出了逗点音符[1]的疑问. 他声称一个完整音符是不可能的，因为它必需把半音符无限地加以逗点. 批评家厄毕米提斯(Epimenides)提出了著名的自相矛盾命题："一切批评家都是说谎者."厄毕米提斯悖论及其现代变种"这支歌是不能唱的"，最终归结为哥德尔(Gödel)的"宇宙交响乐"，它要末不能批评，要末是不完全的. 即便如此，人们对欧几里得和声法的普遍有效性，仍未提出疑问.

非欧几里得

这些不协和音的最终解决者是将近二千年后出生的非欧几里得(1874—1951). 1921 年 7 月末的一天，非欧几里得告诉他的一个学生："今天下午，在 1 点钟差 12 分时，我作出了一个伟大发现，它足以保证在未来 100 年中，德国人在几何学上至高无上的地位."按照这句名言的前面两个单词，他把他的新系统说成是"不成调的"[2]，结果，其新几何果真被人称为"不入调的几何学".

这种"差 12 分就到 1 点钟的几何学"[3]把平行间隔的概念完全打发掉了，从而一劳永逸地排除了招致许多麻烦的根源，并进而宣布任何区间都不受制约. 非欧几里得把钟面取作他的模型，旋律线由时针与分针来识别. 由于它们恒在中心相交，于是就不可能有什么平行的间隔. 在非欧和声法中，每一个不入调的调子都对应着钟面上的一个数. 通过这种尺度可以建立怪调所成之行(级数). 这些级数可以相加、相减、求总和或者求导数. 最后这种运算(怪调求导法)对精神声

[1] 译者注：在音符后加一点，表示延长二分之一.

[2] 译者注：此处语带双关，既有"不成调"的意思，又有"总是第一把交椅"的含义. 因为"at one"即有"在一点"的意思，而拼读起来又成为"atone"，这正是作者的故弄狡狯之处.

[3] 译者注：指黎曼几何.

学家来说是特别感兴趣的.

非欧和声法推翻了简单比的希腊式理想.变异的怪调需要均等调节的完全对称性,在这一音调系统中音程已不再能用简单整数比来表示,也不能通过降半音来把调式的不规则性硬性拉平.非欧和声法切断了它与调节作用的联系,于是它同现实世界也就没有任何直接联系.

多年以来,新和声系统的确切意义规避了人们的批评.人们的确可以任意发明一套新的公理,然后按照那种方式来作曲.但是,其结果会有任何意义吗?造出来的曲子好听吗?入调与不入调,这是完全对立的东西,如果人们抱住新体系不放,那么欧氏和声法的逻辑地位又将如何?欧氏作曲家们的一切努力难道都是白费的吗?

非欧和声法迫使音乐家们拓宽眼界,更深入地思考音乐与现实世界的关系.非欧几里得业已证明确实存在着与欧氏和声法完全针锋相对的一些和声法,可是它们照样能够维持其内在的无矛盾性.这意味着必需把和声法视作一种抽象游戏,独立于它的描述能力.作曲家们现已开始转移他们的注意力,从按照现有方式作曲转为发现新的作曲方式.

对作曲来说,虽然一切和声法有可能都是同样正确的,可是问题依然存在.那就是,究竟哪一种体系确切地描述了声学实际.精神声学方面的晚近研究工作对我们作了最后的揶揄,人们发现耳朵并不能确切地倾听简单整数比.永恒空间被多种其他因素所扭曲,诸如声音的音量,它在传播中的速率等等.因而即使作为真实声音的描述、欧几里得模式也并不正确.事实上,欧氏和声法只不过是包含一切变调的各种可能性所构成的连续统中的一点而已.不过,就日常目的而言,欧氏和声法足敷应用,因为均等调节的音程是泛音系列音程的高度近似.只是在极端情况下(尤其当调制作用以极高速率出现时),两者的区别才明显.正是由于这样的原因,人们在早期未能认识非欧和声法,也就不足为怪了.

最近的进展

以变调几何为其遵循的模式，许多病态作曲法开始出现．有些人试图作出能用一次以上的办法进行演奏的曲子，以产生具有两种或多种功能的组成．有人则鼓吹利用随机因素来表达乐曲的不定形．然而，这种倾向遇到相当阻力．发展了相对音高这一革命性理论的音乐家阿尔伯特·爱因斯坦(Albert Einstein)说了如下的名言："上帝并不玩骰子"．

十九世纪七十年代，乔治·康托(Georg Cantor)开始研究超声音频理论．在以前，作曲家们认定，凡速度突破上界的一切曲子都是同样听不见的．康托表明并非如此，他所用的概念是2对3的一致以及不可数旋律．他的研究最终导致出现一些充耳欲聋的调子，例如黄鹂等鸣禽与红衣主教都可以唱的钢琴曲调．虽然这些声音几乎覆盖了所有的频率，但它们是无法区别的，利用"芭蕾舞滑步法"，康托也得以证明：小提琴要比钢琴拥有更多音符．

1976年，两位钢琴演奏搭档哈肯与阿沛尔(Haken 与 Appel)演出了长久未弹的"地图四重奏"．它表明，每个国家的赞美诗都只须使用四种声调即可唱诵．他们是通过计算机来证明的，计算机对庞大的排列组合数(基本上可从一个极其简单的音乐概念导出)进行了清点计数．哈肯与阿沛尔把一个关于自然美的根本问题复活了：如果没有人听得懂一个曲子的结构，难道能说它是优美的？谁是计算机音乐的作曲家？以上这些问题的答案依然有待于人们的发现．

在非欧几里得与和声法真伪的讨论方面想要掌握更多信息的读者可以参考《哥德尔，埃歇尔，巴哈》(《Gödel，Escher，Bach：一条永恒的金带》)[1]，该书作者是道格拉斯·霍夫斯塔特(Douglas Hofstadter)，1979年由 Basic Books 出版．书中通过隐晦的譬喻，把音

[1] 译者注：经乐秀成编译，中文本已于1983年由四川人民出版社刊行．

乐理论家科特·哥德尔(Kurt Gödel),画家莫里茨·C·埃歇尔(Maurits Cornelis Escher)以及数学家约翰·S·巴赫(Johann Sebastian Bach)[1]的工作编织在一起.

[1] 译者注：实际情况恰恰相反.哥德尔是数学家,而巴赫是音乐家,作者有意将其颠倒,似在向读者进言:本是游戏文章,何必忒煞认真?

有魔力的立方八面体

洛杉矶市大学

查理·W·特列格(Charles W. Trigg)

一个立方八面体的表面含有八个等边三角形与六个正方形.这些三角形与正方形交替地围绕于它的十二个顶点,如同图1中的正交投影所指出的情况.

这些正方形构成了一个顶点连通型网络,在每个顶点上有两个正方形.这些正方形两两一组地落在平行平面内.处于每一对平行平面中间的是一个中介平面,它包含着一个中介正方形的四个顶点.因此,立方八面体的全部顶点落在9个正方形上,而每个顶点有三个正方形在此相遇.

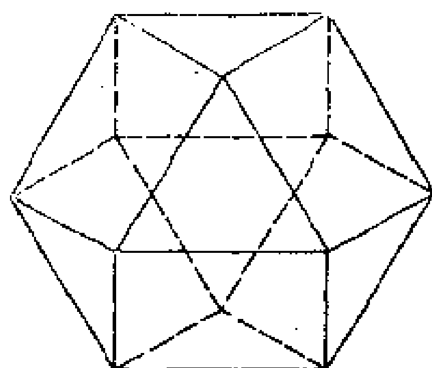


图 1

立方八面体的正交投影.

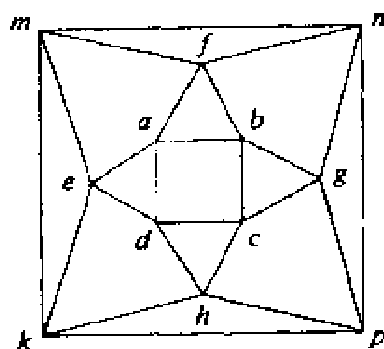


图 2

立方八面体的许莱格尔图❶.

如果总和为 Σ 的十二个元素以这样一种方式分布于立方八面体的十二个顶点,使得每一个表面正方形的周边和数都是 θ 时,则立方八面体称为是***具有魔力的***. 由于周边和的总数满足下式: $6\theta = 2\Sigma$, 故知魔力常数 $\theta = \Sigma/3$. 再考虑三个平行的正方形,可以得出每个中介正方形的周边和数也是 $\Sigma - 2(\Sigma/3) = \Sigma/3$. 在图 2 的许莱格尔图中,中介正方形的四个顶点是 e, f, g, h; a, c, p, m 与 b, d, k, n.

八个表面三角形两两一组地落在互相平行的平面上. 把环绕三角形 afb 与 hpk 的三个正方形的周边和数写成等式并加以化简,即可得出

$$a + f + b = h + p + k.$$

这就说明了,在一个有魔力的立方八面体上,相对三角形的周边和数是相等的.

如果各顶点上的元素即是前十二个正整数,则 $\Sigma = 78$, $\theta = 26$, 作为鉴定魔立方八面体的第一步,下面给出了把 26 分为四个不超过 12 的相异整数的 33 种分拆法. 为了行文简洁,节省地位起见,整数

❶ 译者注: 关于许莱格尔图,读者可参看英国 S. M. P 新数学教科书的有关章节(有中译本).

10,11,12 将分别用 x,y,z 来代表.

12yz	14xy	239z	258y	347z	349x	456y
13xz	159y	248z	267y	356z	358x	457x
149z	168y	257z	259x	348y	367x	4589
158z	169x	23xy	268x	357y	3689	4679
167z	178x	249y	2789			5678

如果把 1 指派到一个固定位置 a,则只有 1 这个数字是公有的两个四数组必须分布在 a,b,c,d 及 a,e,m,f 的位置上. 共有十一对四数组能满足这个要求,它们是:

12yz,	169x;	12yz,	178x;	13xz,	159y;
13xz,	168y;	149z,	168y;	149z,	178x;
14xy,	158z;	14xy,	167z;	158z,	169x;
159y,	167z;	159y,	178x.		

把 1 固定在其位置上,则每个四数组在四边形顶点上的分布方式共有 3! 种,因此共有 $6^2 \cdot 11$,或者说,396 种初始分布需要考虑.

第一个四数组的一种分布方式表示在图 3 的左边. 包含 x,但与

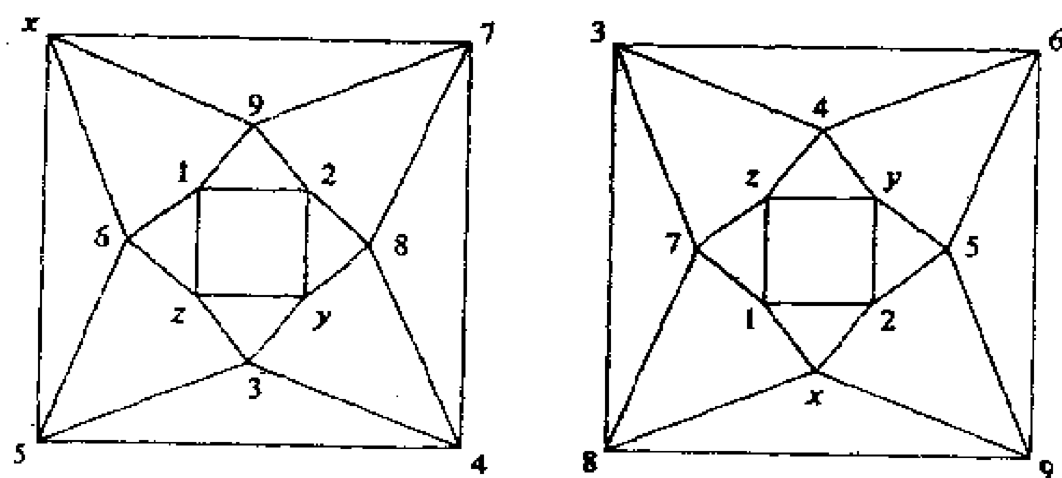


图 3

互补的魔立方八面体.

其他四数组没有同一整数的一个四数组 $457x$ 被选作外层正方形的分布. 这样一来,包含 9 与 2 以及 4,5,7 三个数中的一个的四数组需要安置在 $fngh$ 四边形上,于是,就选一个四数组 2789. 下一步,需要挑选包含 8 与 y 以及 4,5 中两者之一的四数组. 在 $348y$ 被安放在四边形 $cyph$ 上以后,剩下来的数 5 就非放在顶点 k 上不可,这样就最终完成了魔立方八面体各顶点上数的分配.

按照上述办法,已经鉴定了 40 种魔立方八面体,它们可以看作 20 对孪生物,通过由顶点 m、a、c、p 所决定的镜面反射,其中的一个可以转变为另一个. 在表 1 中,只是记下了一对孪生物中的一个. 凡是通过旋转能够重合的分布并不认为是相异的.

如果两个整数的和是 13,则称它们是互补的. 如对应元素都是互补数,则两个魔立方八面体称为是互补的. 图 3 所给出的两个魔立方八面体就是互补的.

在表 1 中,各对互补的魔立方八面体是:$A,B;C,D;E,F;G,H;I,J$. 在每一对中,a,b,c,d 这一组的四个数是同样四个数的重新排列. 对 e,f,g,h 与 k,m,n,p 这两组,除了 A,B 与 I,J 两对之外,情况亦然. 后面提到的两对,四数组则有所交错.

[illegible]	许莱格尔图上的位置												三角和			
	a	b	c	d	e	f	g	h	k	m	n	p	[illegible]	[illegible]	[illegible]	[illegible]
A	1	2	y	[illegible]	6	[illegible]	[illegible]	3	5	x	7	1	20	[illegible]	12	[illegible]
B	1	z	y	2	x	7	4	5	9	8	3	6	18	13	[illegible]	22
C	1	3	x	[illegible]	5	y	[illegible]	2	7	8	4	6	24	18	[illegible]	[illegible]
D	1	z	x	3	y	8	2	5	7	6	4	9	18	15	21	[illegible]
E	1	9	4	[illegible]	7	8	[illegible]	6	2	z	3	y	21	20	18	[illegible]
F	1	z	[illegible]	9	8	6	[illegible]	[illegible]	2	y	3	x	20	18	[illegible]	[illegible]
G	1	5	8	[illegible]	4	x	[illegible]	3	7	y	2	6	23	17	16	22
H	[illegible]	z	8	5	x	9	[illegible]	4	7	6	2	y	17	16	22	23
I	1	7	6	[illegible]	4	y	3	8	2	[illegible]	5	9	26	17	19	16
J	1	z	6	7	5	9	2	x	4	y	3	8	23	13	22	20

（续表）

五角星形	许莱格尔图上的位置												三角和			
	a	b	c	d	e	f	g	h	k	m	n	p	fmn	gmp	hkp	ekm
K	1	2	z	y	6	9	7	4	5	x	8	3	27	18	12	21
L	1	y	z	2	x	8	3	5	9	7	4	6	19	13	20	26
M	1	9	z	4	8	x	5	3	y	7	2	6	19	13	20	26
N	1	y	4	x	6	7	5	8	2	z	3	9	22	17	19	20
P	1	y	4	x	8	5	7	6	2	z	3	9	20	19	17	22
Q	1	8	z	5	6	9	7	4	y	x	2	3	21	12	18	27
R	1	9	y	5	7	6	3	x	4	z	8	2	26	13	16	23
S	1	9	y	5	x	3	6	7	4	z	8	2	23	16	13	26
T	1	y	5	9	7	x	3	6	4	8	2	z	20	17	22	19
U	1	y	5	9	x	7	6	3	4	8	2	z	17	20	19	22

表 1

魔立方八面体，魔幻常数＝26.

从 K 到 U 的十个魔立方八面体是自身互补的. 在 N,P；R,S；与 T,U 这几对中，a,b,c,d 与 k,m,n,p 的分布是一样的. 在 N,P 与 R,S 这两对中，e,f,g,h 的分布正好是互相颠倒过来. 在 R,S,T,U 这一集合中，a,b,c,d 这一四数组表现为同样整数的重新排列，且 e,f,g,h 与 k,m,n,p 这两组也是如此.

由于对应的三角形是正好相对的，因此图 2 中沿着外层正方形境界的诸三角形的周边和数与沿着内层正方形境界的各三角形的周边和数是相同的. 从顶上的三角形出发，按顺时针方向进行，各外围三角形的周边和数都已记录在表 1 之中. 对每一种情形，四个和数都是相异的. 在 C,D；E,F；G,H；K,Q；L,M；R,S 各对中，四个和数都一样，只是排列顺序的不同；在 N,P,T,U 中，情况亦复如此. 在 C,D，E,F 中这四个和数成一等差数列. 在 C,D,E,F,G,H,K,L,M,Q 中，两个和数相间地加起来等于 39；而在 N,P,R,S,T,U 中，两个相继的

和数加起来等于 39.

立方八面体 L, M, R, S 具有超级魔幻性,因为它们的 9 个正方形与 2 个三角形的周边和数均为 26.

在前 12 个整数的所有分配办法中,没有一个能使八个三角形的周边和数统统相等,这是由于 2(78)/8 即 39/2 不是一个整数之故.



● 滑铁卢大学

罗斯·亨斯伯格(Ross Honsberger)

膨胀的铁轨

这则小题目来自默里·克兰金(Murray Klamkin),我们以此作为引子来测试一下马丁的数学直观能力.设想一段平铺伸直的铁轨 AB,长5000英尺,牢牢地固定于两端(图1).在炎热的夏季,铁轨因受热而伸长2英尺,形状变弯了.假如各处变弯的程度是均匀而对称的.那么你猜猜看,铁轨中心点距地面的高度将是多少?一英寸?四英尺,还是只有一英寸的十分之一?

由于铁轨总长度在受热膨胀后将是5002英尺,它的一半便是2501英尺.有鉴于铁轨的弯曲形状将是一条光滑曲线,我们不妨近似地将它视为直线.由此,利用勾股定理,可以算出高 x 的近似值:

$$
\begin{aligned}
x &= \sqrt{2501^2 - 2500^2} \\
&= \sqrt{(2501 - 2500)(2501 + 2500)} \\
&= \sqrt{5001}.
\end{aligned}
$$

由此可见,高度近似等于70英尺!这件事使许多人大吃一惊.当然,这是照直线情况计算的.如把铁轨的弯曲因素也考虑进去,其数值仍

然接近于 67 英尺. 马丁，你想象中的结果，与它接近到何种程度呢？

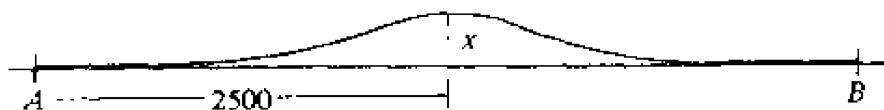


图 1

算术大师

大致一年以前，一种名为“算术大师”的数学游戏风靡我国. 在这种智力玩具中，一位局中人秘密地把四根涂有颜色的木柱排成一列，另一位局中人则企图用最少的步数，结合猜想与推理，来判明木柱的颜色与排列顺序. 下面，让我们也来玩一下类似的算术游戏，它也是默里·克兰金告诉我的.

首先，请你任意选定五个正整数，其中每个数都小于 100，也允许有重复，让我们把它们记为

$$(a_1, a_2, a_3, a_4, a_5).$$

对于这些数字，我当然是一无所知，现在要求我通过下面与你的对话，迅速地判明这些数.

我得向你提交 5 个数 $(x_1, x_2, x_3, x_4, x_5)$，你必须作出回答，不过，我并不要求你像“算术大师”这种智力玩具里的密码编造家那样，要告诉我那些 x 中有多少是正确无误的，有多少是排在正确的位置上的，而只要求你告诉我一个总的和数 S_1 就行了. 这个 S_1 的计算方法是把对应数字相乘以后再求和，即

$$S_1 = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5.$$

当我知道了 S_1，用以武装自己的头脑之后，便向你提交另外一组 5 个数 $(y_1, y_2, y_3, y_4, y_5)$，如果它就是你当初选定的 5 个，那么，我猜对了，游戏到此结束. 如果不属于此种情况，那么你就再告诉我和数

S_2，这个 S_2 由下式决定：

$$S_2 = a_1y_1 + a_2y_2 + a_3y_3 + a_4y_4 + a_5y_5.$$

以上过程就类似地反复进行下去，直到我完全猜对为止.

很明显，只要经过五步，我手头就有五个未知数 a_1, a_2, a_3, a_4, a_5 的五个方程，从而可以把它们求出来. 因此，这个游戏绝对不需要超出五步.

当然，我的主要目的是步数越少越好. 马丁，你相信不相信，我只要用四步就能猜对？用三步就能办得到？二步也行？老实告诉你，事实上只要一步就够了！

由于我知道你选定的任一数字都小于 100，因此我内定的第一组数字是($10^8, 10^6, 10^4, 10^2, 1$)，这必将导致你去计算

$$S_1 = 100000000a_5 + 1000000a_4 + 10000a_3 + 100a_2 + a_1,$$

而其结果如果自左至右，两位撇开以后就将把你所选定的五个数字暴露无遗. 例如

$$(a_1, a_2, a_3, a_4, a_5) = (17, 68, 5, 42, 8)$$

则 $S_1 = 1768054208$，就是那么回事.

前 n 个正整数

人们熟知并易于建立如下等式

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

然而布赖汉姆·扬格(Brigham Young)大学的唐纳德·斯诺(Donald Snow)却采用一种与众不同的办法来得出这一结果.

设

$$S_1(n) = 1 + 2 + \cdots + n,$$

$$S_2(n) = 1^2 + 2^2 + \cdots + n^2.$$

则

$$S_2(n+m) = 1^2 + 2^2 + \cdots + (n+m)^2$$

$$=(1^2+2^2+\cdots+n^2)+(n+1)^2+(n+2)^2+\cdots+(n+m)^2$$
$$=S_2(n)+(n^2+2n+1^2)+(n^2+4n+2^2)+\cdots$$
$$+(n^2+2mn+m^2)$$
$$=S_2(n)+mn^2+2n(1+2+\cdots+m)+(1^2+2^2+\cdots+m^2)$$
$$=S_2(n)+mn^2+2nS_1(m)+S_2(m).$$

即 $$S_2(n+m)=S_2(n)+S_2(m)+2nS_1(m)+mn^2.$$

互换 n 与 m，则有

$$S_2(m+n)=S_2(m)+S_2(n)+2mS_1(n)+nm^2.$$

由于等式 $S_2(n+m)=S_2(m+n)$，上式可立即化简为

$$2nS_1(m)+mn^2=2mS_1(n)+nm^2.$$

由于 $S_1(1)=1$，令 $m=1$，代入后得出

$$2n+n^2=2S_1(n)+n,$$

于是我们的公式 $S_1(n)=\dfrac{n^2+n}{2}$ 就砰地一声爆了出来！

对于下面的特殊关系式

$$1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2,$$

我从未平息过自己的惊讶之感. 在访问澳大利亚时，罗杰·埃格莱顿(Roger Eggleton)给出了这一关系式的几何解释.

由于 k 个边长为 k 的正方形面积之和是 $k\cdot k^2=k^3$，让我们在一个平面上作出如下配置：1 个边长为 1 的正方形，2 个边长为 2 的正方形，3 个边长为 3 的正方形，……，n 个边长为 n 的正方形，则这些正方形的面积总和就是 $1^3+2^3+3^3+\cdots+n^3$. 我们将力图在这些正方形的外面构筑起一个大的正方形，把给定的这些正方形全部包罗在内(见图 2). 容易看到，这件任务基本上完成得很好，可是对于偶数边长 2，4，6，…的那些正方形来说，却是既有重叠(图中的阴影部分)，又有空缺(图中含字母 x 的部分).

稍一迟疑即可恍然大悟，对同样大小的正方形来说，重叠的那个正方形与空缺的那个正方形正好相等. 这样就得出了结论：总面积等于边长为 $(1+2+\cdots+n)$ 的正方形之面积.

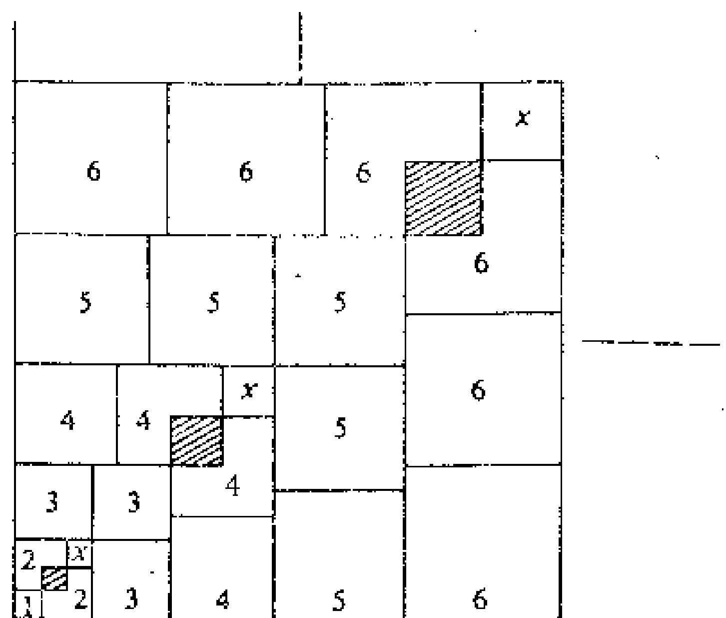


图 2

赫柏·向克(Herb Shank)的送牛奶路线

1974年年中，我的同事赫柏·向克在图论领域中得出一个颇不寻常的发现，我将叙述他的结果但略去其证明. 为了避免卷入对一些图论专门术语作出烦琐解释，让我们把自己局限在通常的几何圈子里. 这样做，定理将不以其最一般的形式出现. 为了说明整个故事，请

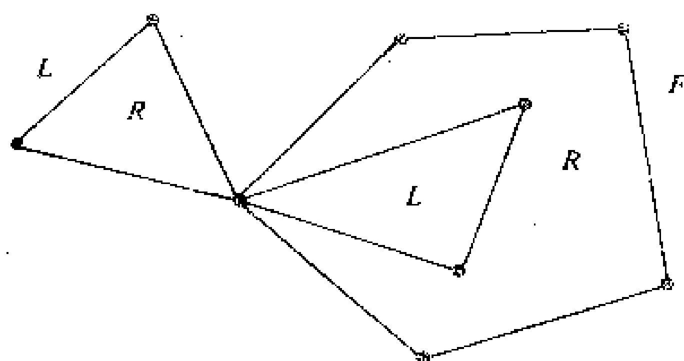


图 3

允许我对熟悉简单图论知识的读者使用下列简单语句：

我们的故事将从一个具有奇数个生成树的任意平面欧拉图开始.

设有一个平面构形 F，它是通过一点而把各个多边形联结起来.你可以随心所欲地使用任意个多边形，甚至一个多边形可以位于另一多边形的内部.我们只需注意下列三项要求必须得到满足：

1. 任一多边形必须具有奇数条边，

2. 任意两条边都不准交错，

3. 所有的多边形都不允许在一个以上的顶点互相联结(这样可以避免产生多边形的"环圈").

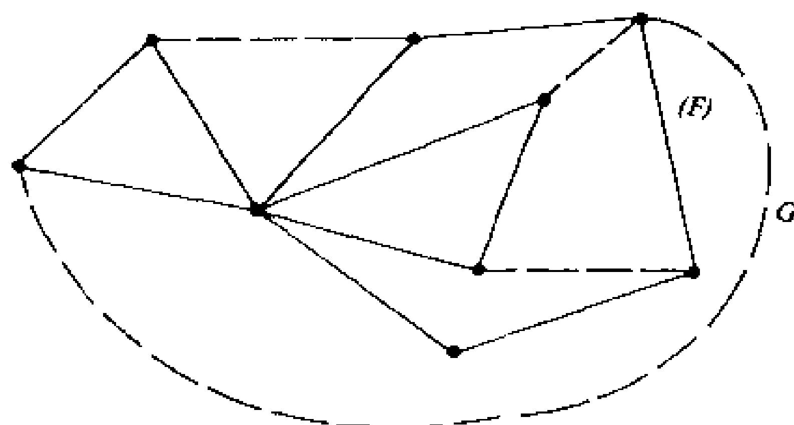


图 4

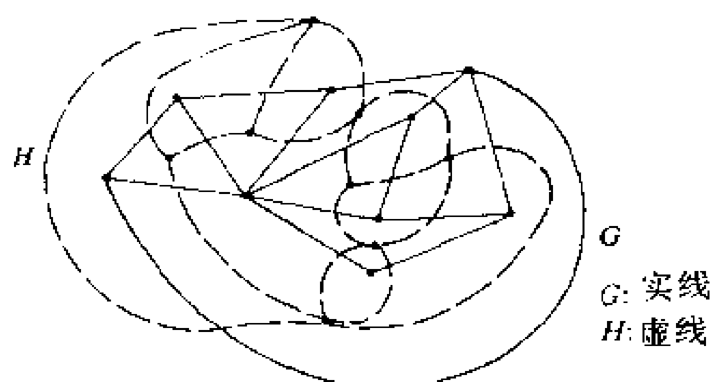


图 5

现在,把平面图 F 的各个区域进行着色,交替地使用红色与草绿色,其原则是:具有一条共同边界的两个区域应当具有不同颜色.上面的三项要求将能保证这样一种着色方案是永远做得到的.为了使我们的例题尽可能保持简单,让我们用两个三角形与一个五边形来说明(见图 3).让我们分别用字母 R 代表红色,L 代表草绿色.

然后,将要通过添加一些附加的边来修饰原来的图形.对 F 中的任一对顶点均可作附加边,只要满足下列条件:任意两条边都不准自身交错或者与 F 中原有的边交错(为此,甚至不惜引入复边,即在已有一条边相连接的两个顶点之间再画一条边).经过修饰后的图叫做 G(图 4).

通过下列步骤,由 G 导出一个独立的图形 H,称为对偶图:

1. 在 G 的每一个区域内标出一个顶点,无限大的外部区域也是如此;

2. 对 G 的两个相邻区域的顶点用一条边相联,使之穿过区域的共同边界;换言之,对构成共同边界的任一条边都应穿插一条边(虚线)(见图 5).

最后,我们对图 H 的每个顶点都根据其所属区域的颜色红或草绿而分别标以 R 和 L(图 6).

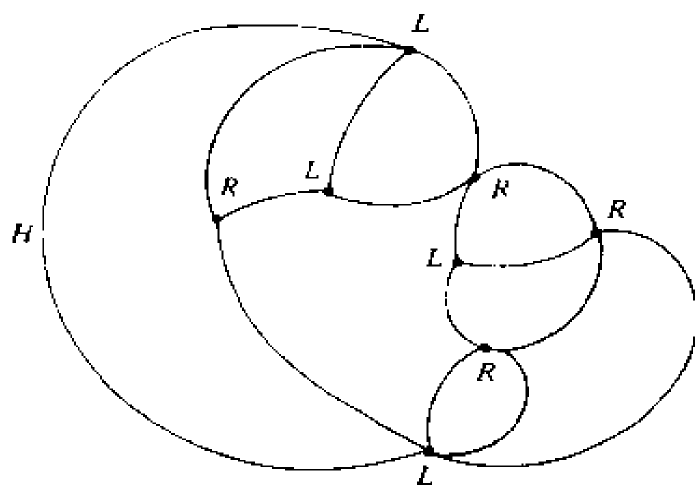


图 6

现在,我们有了一个带有标号的图 H,不管在构作此图时经历了什么不同的选择,它必然具有下面的显著特性:

从图中任一顶点 V 出发,以它为起点,沿着任一条边行进,你定将发现,当重新回到 V 时,每条边都将走过两次,而且是每个方向都各一次.另外,在标有 R 的顶点处必为右转弯,在标有 L 的顶点处必为左转弯.

这样的一种巡回路线可以通过标有数字箭头的有向图加以说明[1].譬如说,我们可以得到下面的图形——图 7.

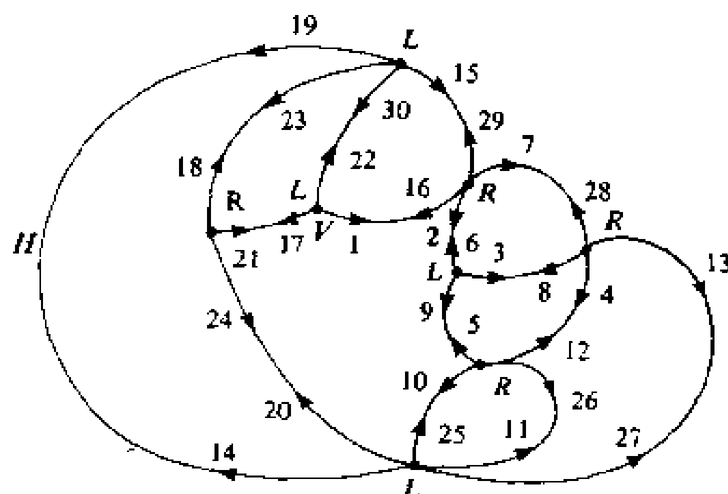


图 7

谁偷了苹果?

我想,我们中间的许多人都曾在度假时做过一些诸如"某人从事什么职业"或者"谁同谁结婚"之类的动脑筋题目.在作这类消遣时,人们所使用的唯一手段,似乎只有一种直截了当的试探消去法.然而,有些问题却有着非常巧妙的解法.例如下面的问题:

在 6 个男孩中,已知有 2 人偷了苹果,但究竟是谁偷的

---

[1] 译者注:数字表示行进的先后顺序.

呢？哈利说："小偷是查理与乔治."詹姆士说："是唐纳德与汤姆."唐纳德却说是"汤姆与查理".乔治的说法是"哈利与查理"，查理说是"唐纳德与詹姆士".汤姆避开了，找不到他，因此也不知道他要说些什么.在四个孩子的回答中都正确地指明了一个人是小偷，而另一个人是不对的.然而第五个❶孩子的回答则是彻头彻尾的撒谎.请问，究竟谁偷了苹果？

本题不无诱人之处：汤姆找不到了，有一个人说谎.在我所写的《数学中的巧智》一书中对此问题给了一个比较复杂的解法.近来，我的好朋友与同事司各特·范士通(Scott Vanstone)告诉我下面的巧妙解法，它把整个问题转化为一个简单的图论问题.利用约翰·阿纽列斯(John Annulis)(位于 Monticello 市的阿肯萨斯大学)的观察，问题的解法简直可以说是清澈见底的.

设有一个图(图 8)，图上每个顶点表示一位男孩(C 代表查理，G 代表乔治，以此类推)，被提到名字的一对男孩，其相应顶点用一条边

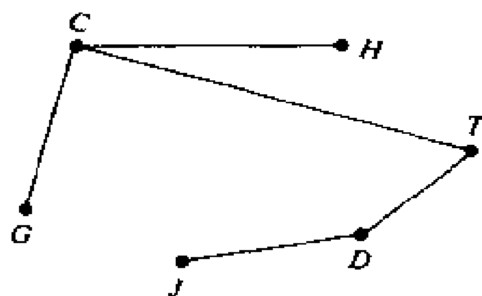


图 8

相联(例如 CG 边反映哈利的话，如此等等).这样一来，除了一条边之外，图中所有各边都有一个端点指向小偷，而那条例外的边则由于反映的是存心撒谎者的话，所以没有任何端点指明小偷.这就是说，这两个顶点，一共有 4 次被作为端点来考虑.(四条边中，每边各有 1

❶ 译者注：这里所谓的第五个孩子并不是按说话顺序所数到的第五人.

个，第五条边则一个都没有）．我们还应注意到，表示两个小偷的顶点之间是不应当有线段相连接的❶．于是从图中立即可以看到，小偷只能是查理与詹姆士（C 与 J）．

克伐塔尔（Chvátal）的艺术画廊定理

在我的书《数学瑰宝 II》中，第十一章的标题是"克伐塔尔的艺术画廊定理"，它涉及到一个画廊中名画的防窃问题．我们知道，博物馆与画廊中的展室，布置得总是蜿蜒曲折，有着各式各样的壁龛与角落．要在每个死角里都有一只眼睛牢牢盯着，确非易事．本问题就是要决定出监视整个建筑物所必需的最低限度的保卫人员．这些监视者应安插在固定位置，但他们可以在那里自由转身．画廊各处的墙壁假定都为直线形状．克伐塔尔证明：如果画廊有 n 垛墙壁（其建筑平面图是图 9 那样的 n 边形），那么，不论其形状何等不规则，最低限度的警卫人员不需要超过$[\frac{n}{3}]$，也就是$\frac{n}{3}$的整数部分．在《数学瑰宝 II》一书中有他的证明．现在我想提出这一定理的一种新证法．克伐塔尔的分析无疑是属于第一流的，但是，以下由波多温（Bowdoin）学

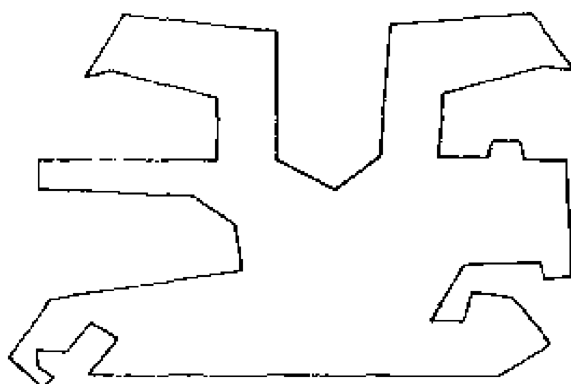


图 9

❶ 译者注：两个小偷之间不应有短线相连，否则将意味着总有一个孩子讲了真话，而不是半对半错了．

院的斯蒂弗・斐斯克(Steve Fisk)的证法则要简单得多.

首先,在画廊内部引一些互不相交的对角线,将它分成若干个三角形(图 10). 其次,把画廊平面图的各个顶点分别着色,其原则是,图中由一条边相连的两个端点必须使用不同颜色. 只须简单地运用数学归纳法即可证明三种颜色即已够用. 因为,对 $n=3$,整个图形仅不过是一个简单的三角形,显然三种颜色是足够的. 现在假设对一个已经分划成若干个三角形的、有着 n 条边的画廊($n\geqslant 3$),三色已经够用,而我们手头的一个画廊 G 却具有 $n+1$ 条边.

显然,总是可以引一条对角线(AB),使它能从 G 中割出一个三角形(ABC)来. 根据归纳法假设,G 中去掉三角形 ABC 所剩下来的,已经三角化了的图形 G' 必可分别用三种颜色来着色,使之满足任意两个相邻顶点具有不同颜色的要求. 由于在 A、B 两点只使用了三种颜色中的两种,下余的一种颜色必可用于 C 的着色. 这样一来,就成功地证明了三色方案定可推广及于三角化的画廊 G 本身. 于是,按照数学归纳法,在一切场合下,三色都够用了. 证明完毕.

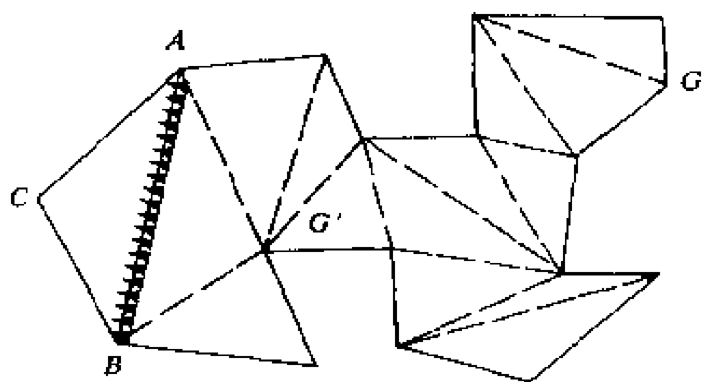


图 10

设想我们的画廊的所有顶点都已用三种颜色 a,b,c 适当地分别着色完毕,既然总共只有 n 个顶点,所以任何一种颜色的使用都不可能超过 $\frac{n}{3}$ 次,其中用得最少的一种颜色(不妨假定是 b),只能出现 m

次，这里 $m \leqslant \frac{n}{3}$. 既然 m 是一个整数，这就等于是说 $m \leqslant \frac{n}{3}$.

既然图上任意一个三角形的两个顶点不可能同色，所以每个三角形的顶点都必须是 a,b,c 三色皆备. 显然，在涂有 b 色的顶点处的警卫员从那里能洞察整个三角形的动静，于是，在着上 b 色的 m 个顶点处的 m 个警卫员就能察看所有的三角形，也就等于是说，控制整个画廊.

驰　　道

现在请看下面引人注目的问题，据我所知，它来自杰出的匈牙利青年数学家拉兹洛·洛瓦兹(Laszlo Lovász).

在一个环形驰道上任意设置着一些加油站 $x_1, x_2, \cdots, x_n$，每个加油站都储备着一定数量的汽油，其全部数量恰巧可供一辆汽车沿着驰道 T 跑一圈(图 11). 请证明，不管汽油在各加油站之间如何分配，

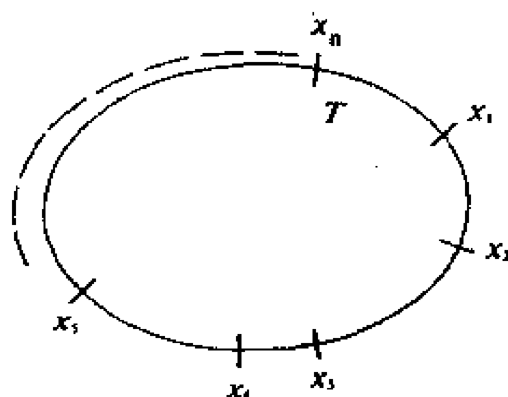


图　11

总存在着那么一个加油站，使得一辆汽车可以在那里起跑，然后沿途加油，循着整个驰道正好跑一圈.

下述解法由迪安·霍夫曼(Dean Hoffman)(Auburn 大学)提供，他是从洛瓦兹本人那里直接学来的.

设想一辆汽车带足额外汽油，从驰道上的任意一点出发作试跑.

由于所需要的汽油总量正好等于在沿途各加油站所供给的汽油总量，因此在一圈结束时，汽车的剩油量恰好等于开始时所带的汽油量. 现在，让我们在汽车行驶过程中紧紧盯住油量仪表，我们注意到当汽车抵达加油站 x_i 时，仪表读数在整个行程中最低，设其时的读数是 d 加仑. 因此，在整个行驶过程中，我们总是带了过剩的 d 加仑油，这些油量是从来不用的. 很明显，如果把这 d 加仑油放在家里不带出去，一切都将照常进行. 不带这些不需要的油，我们将会在油箱开始空空如也的时候正好赶到 x_i 这一加油站，而这一情况同我们在加油站 x_i 开始起跑是完全一样的，以 x_i 为起点，油量计的读数将永远不至于在 0 点之下，因为在试跑过程中它的读数永不会在 d 加仑之下.

格点立方体

三个坐标都是整数的三维空间的点 (x,y,z) 称为格点. 这种情况将发生于用一个单位立方体的各个顶点去度量被平面 $x=a, y=b, z=c$ 所分割的空间，此处 a,b,c 都是整数. 显然，存在着无限多立方体，它们的 8 个顶点全是格点，而且其各个面都平行于坐标平面. 这种立方体的边长显然是一个整数.

还有许多立方体，它们的顶点虽为格点，可是却没有摆在“标准”位置，而是歪斜地坐落着(对格子的框架而言). 我们中间大概很少有人会具备这种“先见之明”，对下列包罗一切的结论毫不感到惊讶的. 这一结论如下：

所有顶点都是格点的立方体 C，其边长 s 恒为整数.

由于格子中究竟取哪一点作为坐标原点都属无关紧要，于是我们可取 C 的任一点作为原点 O. 设从 O 发射出去的三条边止于顶点 $V_1(x_1,y_1,z_1)$，$V_2(x_2,y_2,z_2)$，$V_3(x_3,y_3,z_3)$，由于边长是 s，所以立方体 C 的体积 $V=s^3$. 有一个大学低年级学生熟知的公式，它把立方体的体积通过顶点坐标来表出，如适当选取符号，即有

$$V=\pm\begin{vmatrix}x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\\ x_3 & y_3 & z_3\end{vmatrix}.$$

由于这些数全是整数，所以体积 V 也必须是整数，

即

$$s^3\text{ 是一整数.}$$

可是边 OV_1 之长是

$$s=|OV_1|=\sqrt{x_1^2+y_1^2+z_1^2}=\sqrt{\text{一个整数}}.$$

于是可知 s^2 也是一个整数. 这样，s 本身就非有理数不可：

$$s=\frac{s^3}{s^2}=\text{有理数}.$$

但正如我们刚才所看到的，$s=\sqrt{\text{一个整数}}$. 如果一个整数的平方根不是无理数的话，那它不仅是有理数，而且必是整数. 由此可见，s 是一整数.

以上结果可用一种平铺直叙的方式推广到 n 维空间的超立方体，这里的 n 是奇数.



● 斯坦福大学

□ 威廉・T・拉塞尔(William T. Laaser)

□ 莱尔・雷姆肖(Lyle Ramshaw)

摘　　要

本文研究了转台问题的下列推广：台子的形状是一个正 n 边形，在它的每一只角上有着一个凹坑，里面有一只直立或者倒放的玻璃杯. 游戏者具有 k 个探手. 我们的问题是：对于任意的 k 与 n，是否存在着一种办法，使得经过有限步之后，足以保证铃声大作. 在这篇文章里，我们将会看到，当且仅当参数 k 与 n 满足下列不等式

$$k \geqslant \left(1-\frac{1}{p}\right)n \qquad (p\text{ 是 }n\text{ 的最大素因子})$$

时，这种办法确实存在.

第一节
引　　言

在 1979 年 2 月号《科学美国人》数学游戏专栏里，马丁・加德纳

提出了下列转台问题：

我要开始讲一个来路不明而非常讨人喜欢的新鲜组合问题，它是多伦多的罗伯特·塔沛(Robert Tappay)告诉我的，他相信这个问题来自苏联[1].

设有一只方桌子，它可以绕着其中心转动.在每只角上有一个很深的凹坑，其底部放着一只直立或倒立的玻璃杯.不准你用眼睛去看，但是你可以用手去触摸酒杯以弄清楚它们的放置方式.

游戏的一步可以定义如下：转一转台子，当它停下来的时候，伸出你的双手，摸进两个不相同的凹坑，你可以调整玻璃杯的摆放方式，高兴怎么做就怎么做.也就是说，你可以保持玻璃杯的原状，改变一只或两只玻璃杯的摆法.

现在，你可以转动台子，重复以上过程，作为你的第二步动作.当台子停止转动时，是没有办法区别它的每只角的，因而你只可能有两种选择：要末是你用双手摸进互为对角线的两只凹坑，要末是相邻的两只.游戏的目标是要使四只玻璃杯按同样方式放置：或者全是正放，或者全是倒立，一旦做到了这一点，铃声就响起来.

开始时，四只凹坑里的玻璃杯是任意放置的，如果它们的上下摆法恰巧相同，那么铃声就会顿时响起，什么步骤都不必采取，游戏任务即已宣告完成.因此，我们可以假定，在开始时，玻璃杯并不是按同样方式摆放的.

有没有一种办法，足以保证在有限步数内必可使得铃声响起来.在他们粗略地思考了一下这个问题之后，许多人下结论说这种办法是不存在的.他们的根据是，它是一个概率问题.如果运气不佳，人们将会无止境地重复其步数而毫无所获.然而，实际并不如此.在不超过 n 步正确动作之后，

[1] 译者注：在俄文文献中未找到此问题.

人们一定有把握使铃声震响. 那么,n 的最小值应是多少? 要采取何种步骤以使得铃声在 n 步或更少的步数内响起来?

设有一张仅有两只角的桌子,因而它一共也只有两个凹坑. 在此种情况下,一步动作即足以使铃声大作. 如果有三个凹坑(它们位于正三角形桌子的三个角上),则下面的两步动作即已足够:

1. 用手探查任一对凹坑,如果两只玻璃杯是按同样方式摆放的,那就都加以颠倒,这时就会响起铃声. 如果它们的摆放方式不同,则把那只面朝下的玻璃杯予以颠倒,如果此时铃声不响,则可继续做第二步.

2. 转动台子,并用手探查任一对凹坑,如果两只杯子都是朝上的,则把它们都加以颠倒,此时铃声就会响起来. 如果它们摆法不同,则把面朝下的那只杯子颠倒,则铃声就会响起来.

虽然有着 4 只凹坑与 4 只玻璃杯的本问题可以在有限步数内得到解决. 然而,如有 5 只或更多的杯子(它们分别位于五边形或边数更多的多边形之角上),却不存在一种办法能够足以保证在 n 步内解决问题. 在下一个月的本专栏上, 我将给出有 4 只杯子的本问题解法, 并将讨论由罗纳德·L·格兰汉姆(RonaldL. Graham)与珀西·迪康尼斯(Persi Diaconis)所作的若干推广.

还有,若干细节有必要加以澄清,即有关游戏者所采取的步骤怎样才算是合法的问题. 首先,在动手之前,游戏者必须明确宣布她要动手摸索的凹坑究竟是哪两个. 如果她先去摸一个坑,然后根据所摸索到的结果再来决定要摸的另一个坑,这样的做法是不允许的. 其次,当她进行了摸索并将玻璃杯的放法作了某种变动以后,在她把双手从凹坑中撤回之前,标志成功的铃声仍然是不会响的. 这就是说,下列动作被认为"不合法":游戏者用她的手在凹坑内把玻璃杯摆来

摆去，企图在至少一种摆法上听到铃声.

在1979年3月号《科学美国人》杂志马丁·加德纳的专栏上，有着四个坑与四只玻璃杯的本问题解法已经给出，最多五步，铃声便响. 如果读者对转台问题尚不熟悉，我们鼓励他自己去重建这个解法.

1979年3月号的专栏也提到了由罗纳德·L·格兰汉姆与珀西·迪康尼斯所建议的转台问题的两种推广. 他们首先建议游戏者可以多于双手[1]. 对于参数 k 与 n，我们可以设想一个具有 k 只触手的游戏者，一张正 n 边形的转台，以及相应的 n 只角、凹坑与玻璃杯. 我们的问题是要发现一种办法，足以保证在有限步之内使铃声鸣响，或者从反面证明不存在这样的办法. 格兰汉姆与迪康尼斯也建议对本问题作另一种方式的推广：不用玻璃杯，而代之以具有两种以上位置的物体. 其他有趣推广法也还是存在的，例如斯科特·金(Scott Kim)曾考虑不用方桌，而改用对称性更为丰富得多的，例如，立方体形状的桌子.

本文专门讨论这些推广中的第一种. 特别，我们将对 $n\geqslant 2$ 定义函数 $f(n)$ 以表示相应的最小正整数 k，使得有 k 只触手的游戏者在面对正 n 边形转桌与 2 状态的玻璃杯时，存在一个有限的过程，足以保证使铃声鸣响. 我们的目标是要决定函数 f 的值.

有关转台问题的一些游戏规则相当微妙，很难说它们已得到了完全确切的说明. 读者们不妨回忆一下，在上面，我们曾经被迫停顿，以澄清若干细节. 它要求我们必须把问题叙述得更形式化，因为，更形式化的描述将是更为准确的. 我们将把本问题重行叙述为一种非对称的两人博弈，其局中人是游戏者与转台，并涉及圆形字符集的处理.

在本文中凡是谈到多边形，我们都是指那种可以围绕其中心进行自由转动的正多边形，但是不能上下翻转. 在这种理解之下，我们

---

[1] 译者注：可多于双手，这当然不是指生理意义. 以下，为区别起见，我们将使用"触手"这一名词.

来定义长度为 n 的圈，它是正 n 边形顶点的一组标号. 长 n 的圈与长 n 的字符串颇为类似，然而圈中字符的圆形轮换都被认为仍然是这个圈. 对于特定的圈，既可用图形也可用括号来标明. 在括号中，可以圈中的任一点作为起点，并按顺时针方向相继记下其他各点. 例如，图 1 中长度为 6 的圈既可记作〈010011〉，也可记作〈100110〉，此外还有另外四种记法. 然而〈101100〉则表示另一不同的圈.

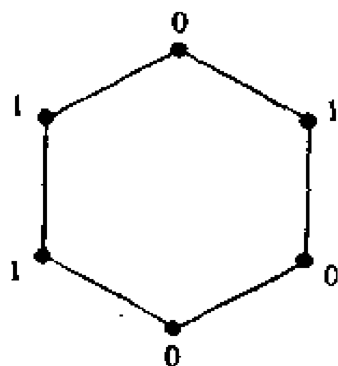


图 1

一个长度为 6 的圈.

在某一时刻，凹坑中玻璃杯的即时状态对应于由字母 $\{u,d\}$ 组成的长度为 n 的一个圈，这里我们以字母 u 与 d 分别表示玻璃杯的正放与倒置状态. 转台游戏的第一步是由局中人"转台"一方所作出的. 特别地，转台一方通过选取 $\{u,d\}$ 上的一个圈作为玻璃杯的初始状态. 我们把这种状态记为 S_1.

现在该轮到游戏者的第一步动作，她必须公开宣布她的试探触手对转台的圆形安排. 我们将称这种安排为探查模式. 形式地讲，探查模式可由字母 $\{h,g\}$ 组成的一个圈来表示，这里 h 表示"触手"，它是游戏者企图触摸的某个位置，g 则表示"空白"，它是游戏者不拟进行触摸的位置. 当然，游戏者选择的探查模式中，符号 h 的出现不应超过 k 次，因她只有 k 只触手. 我们把这第一个探查模式称为 P_1.

接下来又是转台的动作. 它从圈 S_1 与 P_1 重叠的 n 种可能方式中选取其一. 这种选择即意味着转台的旋转. 在转台一方作了这一选

择之后，游戏者即被告知，在探查模式 P_1 中每一个 h 的位置，状态 S_1 中的对应记号究竟是 u 还是 d. 于是，游戏者可将测试位置上的任意一个，或者所有的 u 变作 d，反之也可将 d 变作 u. 这样改变的结果就定义了转台的下一个状态，它是字母 $\{u,d\}$ 上的圈 S_2. 如果一个圈中的所有字母全是 u 或者全是 d，我们就把圈的这种状态称为单纯的. 如果状态 S_2 是单纯的，铃声就会鸣响，而游戏者在一次试探中就赢了.

按这样的方式，游戏可反复进行下去. 设我们刚刚确定了一个非单纯状态 S_i，于是游戏者采用了某一探查模式 P_i；转台决定 S_i 与 P_i 如何重合，游戏者被告知与 P_i 中 h 相对应的 S_i 的位置上的情况，然后游戏者决定这些位置的新情况，由其决策决定了一个新的状态 S_{i+1}. 如果 S_{i+1} 是单纯的，则响声就响，而游戏者被认为在 i 次探查中获胜. 转台一方的目标是阻挠游戏者取胜，如果本游戏永远继续下去、没完没了，我们就认为转台获胜. 请回忆一下，我们是把 $f(n)$ 定义为最小的正整数 k，它能使游戏者通过 k 个触手而获胜. 如果 $k\geqslant f(n)$，则对游戏者一方存在着一个得胜策略；如果 $k<f(n)$，则转台一方存在一个得胜策略.

本文余下部分是专门来证明下列结果的：

定理 1 对 $n\geqslant 2$ 的任一整数，设 $f(n)$ 表示具有下列性质的最小整数，即 $f(n)$ 只触手足以使游戏者赢得转台博弈. 如果 p 是 n 的最大素数因子，则 $f(n)$ 的值将由下式给出

$$f(n)=\left(1-\frac{1}{p}\right)n.$$

前人的工作业已证明了这一结论的某些特殊情形，正如马丁·加德纳在 1979 年 3 月号的专栏中所公布的，罗纳德·L·格兰汉姆与珀西·迪康尼斯证明了，如 n 为素数，则 $f(n)=n-1$；如 n 为合数，则 $f(n)\leqslant n-2$. 对所有的 $a\geqslant 1$, $b\geqslant 2$，斯科特·金证明了下界

$$f(ab)\geqslant a\left\lceil\frac{b}{2}\right\rceil,$$

这里$\lceil x\rceil$表示把 x 取到不小于 x 的最小整数[1]. 詹姆斯・伯伊斯(James Boyce)是猜测到定理 1 中 $f(n)$表达式的第一人.

顺便说一句,定理 1 有一个奇妙的推论:除去 $n=2$ 的情形,所需的最少触手数恒为偶数.

第 二 节
下 界

在本节中,我们将论证定理 1 所给出的公式将构成 $f(n)$值的一个下界. 特别地,我们要证明下述引理:

引理 1 设 n 是一个$\geqslant 2$ 的整数,p 为其最大素因子. 若游戏者只限于使用较 $\left(1-\frac{1}{p}\right)n$ 为小的触手数,则在转台博弈中,台子一方必存在一个取胜策略.

证明 我们首先注意到,若 n 分解为 $n=lm$ 的形式,则一个 n 边形可看作 l 个相异的、相互穿插的 m 边形. 例如,由于 $6=2\times 3$,我们可以把六边形看成是一个 David 星形,如图 2 所示.

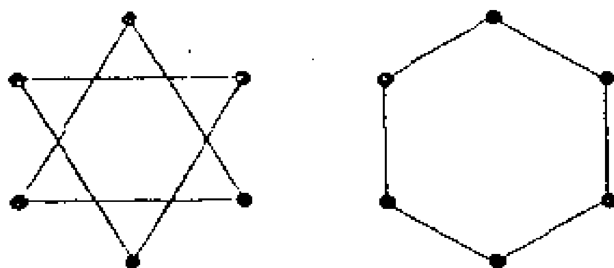


图 2

把六边形看作两个相互穿插的三角形.

设 p 是整数 $n\geqslant 2$ 的最大素数因子,并通过关系式 $n=lp$ 定出了 l. 实际上,在这一论证中我们也可把 p 取作 n 的任意素数因子,但是

---

[1] 译者注:例如$\lceil 3.18\rceil=4$,注意此记号同一般的取整函数$[x]$有异.

最大素数因子将给我们以最强的结果. 如同上文注意到的，转台游戏的状态与探查模式可以看作是由 l 个相互穿插的 p 多边形所组成. 假定游戏者一方的触手数少于 $\left(1-\frac{1}{p}\right)n$，我们的目标是要为台子一方找到一个取胜策略. 基本想法是对一切 i，要保证至少有一个构成状态 S_i 的 p 多边形是非单纯的，也就是说，既有向上的杯子，也有向下的杯子.

由于转台可任意选定状态 S_1，所以开始时很容易使上述条件得到满足，但是我们必须证明，当本游戏继续玩下去时，转台这一方将仍能维持这个条件. 设状态 S_i 包含一个非单纯的 p 多边形. 选定一个，并称之为 S_i^*. 设 P_i 表示游戏者在第 i 步动作中打算要用上的试探模式. 由于游戏者的触手数少于 $\left(1-\frac{1}{p}\right)n=l(p-1)$，所以 P_i 的 l 个相异的 p 多边形中至少有一个多边形必须至少含有两个空白 g. 选取一个这样的 p 多边形，并称之为 P_i^*. 同样也选好出现在 P_i^* 中的两个缺口，并假定它们在围绕 P_i^* 中相隔的距离为 j. 这里顺便提一下，作为我们的距离单位是，凡相邻顶点的距离看作是 1. 现在，让我们考虑每一步走 j 个顶点，经若干步周游 S_i^* 的问题. 由于 p 是一个素数，所以这种巡回方式将在返回其出发点之前，遍历 S_i^* 的每一个顶点，不管 j 的具体数值如何，都将如此. 在巡回过程中，我们肯定会遇到杯子朝上的状态 u 与朝下的状态 d，这是由于 S_i^* 在选定时就是非单纯之故. 因此，我们将会在碰到一个 u 之后立即碰到某个 d. 而这意味着 S_i^* 中将含有相隔距离为 j 的一个 u 与一个 d.

现在我们已作好准备，可以教转台这一方如何进行下去. 它应把圈 S_i 与 P_i 如此结合，以使得 p 多边形 S_i^* 与 P_i^* 互相重迭，并进而使 S_i^* 中距离为 j 的 u 与 d 与 P_i^* 中距离为 j 的两个空白缺口相对应. 由于游戏者无法改变与空白缺口相对应的杯子的状态，因而所得出的状态 S_{i+1} 将会至少包含一个非单纯的 p 多边形. 通过这种手段的反复运用，转台方面将能对所有的 i，都可使状态 S_i 包含一个非单纯的 p 多边形. 这样一来，就构成了转台的取胜策略.

第三节 完全幂次

引理 1 所提到的下界结果只是全部证明中较容易的部分. 为了完成定理 1 的证明,我们还有艰难得多的工作要做,就是替触手数最少的游戏者设计一套取胜策略. 在开始从事这件任务之前,需先加领会:以前为较小的桌子所构筑起来的策略可以作为子程序,而在为更大的桌子设计策略时派上用场.

我们需要一个定义. 假定桌子的大小 n 可以分解因子为 $n=lm$. 请回忆一下,我们可以把一个 n 边形看作 l 个相异的、互相穿插的 m 多边形. 设有某个长度为 n 的圈 T 具有下列性质:构成它的那些 m 边形全都是单纯的. 这样,T 就必须具有 m 个一式一样的拷贝,其中每个拷贝都具有长度 l,而且首尾相接地衔接在一起. 也就是说,对某种串 X,T 必然具有$\langle X^m\rangle$的形状. 我们将把这样性质的圈命名为完全 m 次幂,以区别于其他圈. 例如,一个具有偶数长度的圈可称为完全正方形,如果任一对角线的两端标号都是相同的话,这就意味着,对某个 X 串行来说,此圈具有$\langle XX\rangle$的形状. 其实,把一个长度为 n 的圈说成是一个完全 n 次幂,就等于是说这个圈是单纯的.

设想我们要为游戏者设计一个策略来对付大小为 n(n 可分解成因子,即 $n=l\ m$)的转台,而且我们已处于一种状态,此时转台的当前状态 S 恰好是一个完全 m 次幂,如果我们始终具有至少 $mf(l)$个触手,我们就能以一种很简单的策略进行到底. 其办法是:只要照搬一个足以对付大小为 l 的转桌的游戏策略就行,把一切事情都重复 m 次. 人们可以把此种过程称为一个策略的乘幂. 先取较小策略的第一个试探模式,然后把 m 个拷贝串联起来,就得到较大策略的一个试探模式. 由于转台的当前状态 S 是一个完全 m 次幂,所以我们那位游戏者的 m 组 $f(l)$触手都将会触摸到一式一样的“上”或“下”序列. 下面我们只要教游戏者,对规模较小的转台你是怎么做的,此时也怎

么去做就成了.不管这种调整究竟如何,较大转台的下一状态仍将是一个完全 m 次幂,因此,我们总是可以自始至终地继续照搬对付较小转台的策略.于是,上述论证已经肯定了下列引理的真实性.

引理 2 设想转台游戏的一方是一只大小为 n 的转台,而台子的当前状态是某个能整除 n 的、不小于 2 的整数 m 的完全 m 次幂.则存在着一个游戏者可运用的策略,它将保证他只要利用 $mf(n/m)$ 个触手,在有限次试探中使铃声鸣响.

在我们能够应用引理 2 以前,我们首先需要迫使转台进入一个完全幂次的状态.下一个引理断言确实存在着处理这部分工作的策略而并不需要用上太多的触手.

引理 3 如果 p 是一个不小于 2 的整数 $n\geqslant 2$ 的最大素数因子,则在规模为 n 的转台游戏中存在着一个游戏者的策略,它具有如下性质:至多利用 $(1-1/p)n$ 个触手,在有限步中,要末使铃声震响,要末将迫使转桌进入某个完全 p 次幂的状态.

我们将把引理 3 的证明暂时推迟一下.现在让我们来证明,把引理 1,2,3 结合起来,即可完成定理 1 的证明.

定理 1 的证明 定理 1 断言 $f(n)=(1-1/p)n$,这里 p 是 n 的最大素数因子.由引理 1,我们得知 $f(n)\geqslant(1-1/p)n$,余下来的工作是要证明 $f(n)\leqslant(1-1/p)n$.我们的证法将是对 n 用归纳法.也就是说,假定对一切规模比 n 为小的转台,定理已成立,我们的任务是要证明对游戏者,确实存在着只利用 $(1-1/p)n$ 个触手的取胜策略.开始时,我们可以先应用引理 3 的策略,这时,要末响起了铃声、要末迫使转台进入一个完全 p 次幂的状态 S.如果 n 是素数,则 p 等于 n,而任何一个等于 p 次完全幂的状态也必定是单纯的.于是,在此种情形下,来自引理 3 的策略总是可以使铃声响起,我们成功了.

另一方面,若 n 不是素数,则 p 是 n 的一个非平凡的因数,因而可应用引理 2.来自引理 2 的策略可使游戏者从状态 S 进入使铃声

响鸣的状态. 我们只需要验证一下,这位游戏人有足够多的触手来实现引理 2 所说的策略就行. 而这归结为验证下面的不等式.

1. $$\left(1-\frac{1}{p}\right)n \geqslant pf(n/p).$$

为了看出这一点,我们将对大小为 n/p 的转桌引用归纳法假设. 设 q 是 n/p 的最大素数因子,也就是 n 的第二个最大素因子. 由归纳法假设,我们可以断定

2. $$f(n/p) \leqslant \left(1-\frac{1}{q}\right)\left(\frac{n}{p}\right).$$

由于 $q \leqslant p$,所以不等式 2 蕴含着不等式 1,所以我们的游戏者具有足够的触手来应用引理 2 所说的策略.

本文的余下篇幅将专门用来证明引理 3. 看来,这个证明将分为两种稍有不同的情形. 如果 n 的最大素数因子至少是 3,则运用一个名为上-下策略的较为简单的策略即可解决问题,这将在第四节中予以说明. 在第五节中,我们将叙述 n 为 2 的整数幂时必须采取的,更为巧妙的上-翻策略.

第 四 节
上-下 策 略

在本节中,我们将对一切不是 2 的整数幂的转台来证明引理 3. 为了对一般情形增添一些直观想象,我们将从 $n=12$ 这个特例出发. 12 的最大素数因子是 3. 因此,对 $n=12$,引理 3 断言如下:存在着一个取胜策略,具有 8 只触手的游戏者可用它来对付一只有 12 只角的转台,使铃声震响或者迫使转台进入一个完全立方的状态.

当 $n=12$ 时,转台形状是一个正十二边形. 为了便于理解我们将设计的策略,需将正十二边形看作两个不相连贯的六边形,而每一个六边形又可看作两个互不连贯的三角形. 图 3a 画出了一个正十二边形,它的各个标上号码的顶点就像是一只普通的 12 小时的钟面. 如果使用略有不同的说法,则图 3a 可以记为下述钟形图:

C=((1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12).)

图 3b 则再次重绘了钟形圈 C,其目的是要反映十二边形被分解为六边形与三角形的情况. 图 3b 中每个矩形都表示钟面上的一些次级正多边形. 其顶点被圈在矩形之中,每个矩形的左上角是一个标号,它给出了圈在其中的顶点个数. 我们把这种图形称为结构图. 对任意长度为 12 的圈,我们都能作出类似的结构图,特别地说,在设计我们的取胜策略时所出现的状态与试探模式也能这样做.

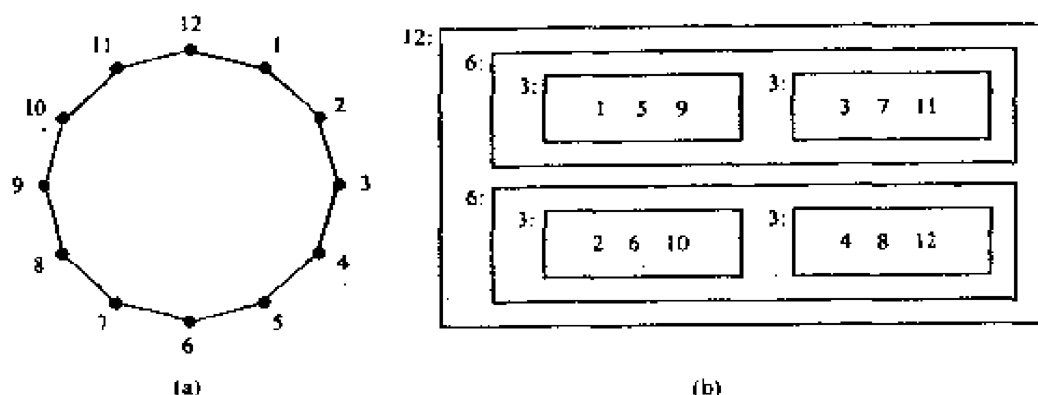


图 3

(a)钟形圈、及其结构图(b).

我们的策略将用上三个不同试探模式 P_1,P_2 与 P_3. 图 4 画出了这些试探模式所应有的结构图. 模式 P_1 是利用 6 只触手确切地试探两只六边形中的一只;模式 P_2 用的也是 6 只触手,但却是在两只六边形中各自确切地试探一只三角形;P_3 用的是 8 只触手,其作用是对四只三角形中的每个三角形确切地试探其中的两个顶点. 不管台子怎样转法,我们所说的这些模式要去探查哪些顶点的盘算都是正确有效的. 我们并不知道在两个六边形中 P_1 倒底要探查哪一个,不过它反正总是要探查其中的一个;对 P_2 与 P_3 来说,情况也是如此. 正是此种转动不变性才使得结构图能派用场.

不管我们画出什么样的结构图,总是至少有一个圈对应于那种结构图,但也可能不止一种. 在 P_1 与 P_2 的情形,恰恰只有一种;如果

我们要有图 4 的结构图的话，就必须选择 $P_1=\langle ghghghghghgh\rangle$ 与 $P_2=\langle gghhgghhgghh\rangle$. 但在 P_3 的场合，我们可以有几种选择. 两个不同的圈

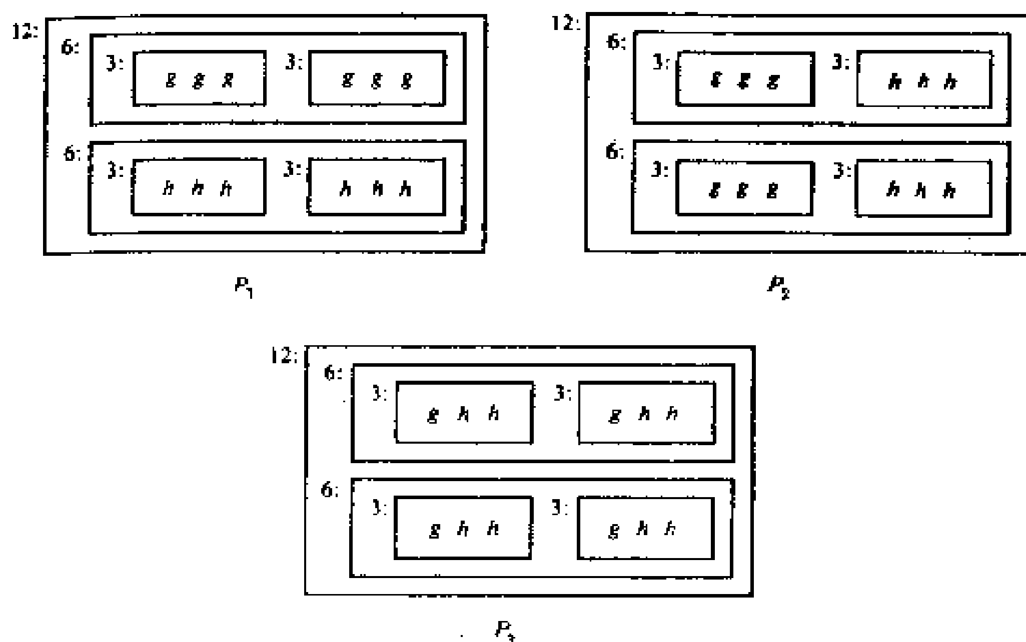


图 4

各个 P_i 的结构图.

$\langle gggghhhhhhhh\rangle$ 与 $\langle ghhghhghhghh\rangle$ 都符合我们要求 P_3 所应具有的结构图；实际上，还有其他六个圈也有这种结构图，不过我们无需特别费心把它们一一列举出来. 结果是，我们可以选取任何一个与正确结构图相对应的圈作为试探模式，至于确切地说，到底要试探哪一个则无关紧要.

在深入讨论这些探查模式之前，我们需要把记法推广，使对状态也适用. 请回忆一下，转台的确切状态由长度为 n 的圈通过字母 $\{u, d\}$ 来表示. 但我们也可同意使用记号 e，以表示转台的部分信息. 记号 e 表示"任意"，它表示在这个位置上既可以杯口向上，也可以杯口向下. 例如，转台 S_1 的初始状态可由公式 $S_1=\langle e^n\rangle$ 给出. 这是由于转台可以任意选择初始状态之故.

作为我们的取胜策略的前三个试探模式，我们可以教游戏者依

计行事，按 P_1，P_2，P_3 的次序进行探查，对她所触摸到的每只杯子，不管它原来的状态是什么，都使之杯口向上．其效果将由图 5 给出，它显示出了状态 S_1 至状态 S_4 的结构图．模式 P_1 的试探将使一个六边形全部变为单纯，即状态 S_2．模式 P_2 的 6 只触手将从每只六边形中试探一个三角形，因此第二次探查后将迫使四个三角形中的第三个三角形也变作单纯，即图中的状态 S_3．而模式 P_3 是要探查每一只三角形中的两个顶点，所以经过第三次探查后又将使留下来的三个 e 中有两个变为杯口向上．总之，在前三次探查后，转台状态 S_4 将变成 $S_4=\langle eu^{11}\rangle$．

也有可能碰巧，在三次试探中，说不定留下来的 e 恰巧都成为杯口向上的状态，于是铃声响起来了，游戏者获得胜利．此种情况，显然不能算游戏者的本领，因此我们可以略去，而把注意力集中于最后一个 e 所代表的是一只杯口向下的状态．因而，状态 S_4 实际上是 $s_4=\langle du^{11}\rangle$．

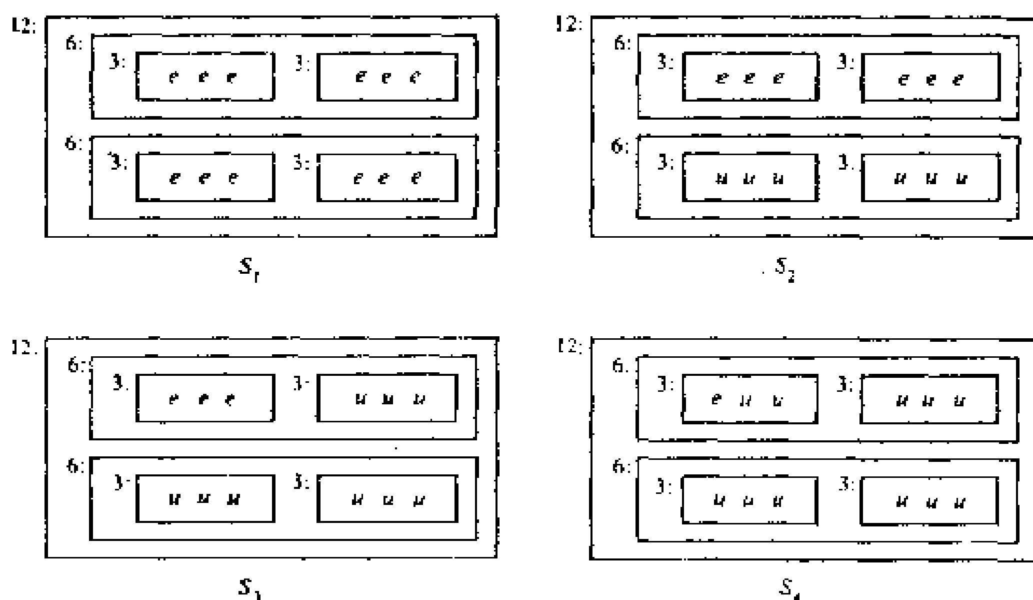


图 5

由 S_1 至 S_4 的结构图．

迄今所讨论的这部分策略我们将称之为第一回合. 第一回合的三次试探，其净效果是迫使转台转入一种只差一只杯子就都是全体杯口向上的状态. 我们将称这只孤单的向下杯子为例外杯子，并把含有它的三角形也称做例外的.

下一步将指向哪里呢？我们的目标是迫使转台进入一个完全立方状态. 至多再进行三次试探即可做到这一点，我们将称之为第二回合的试探. 这些试探同样也按模式 P_1, P_2, P_3 相继地进行. 首先，让我们按 P_1 来探查 S_4；请回忆一下，P_1 是探查两个六边形之一的. 如果台子转过以后，游戏者恰好触摸到了那只例外的杯子，于是他只要简单地把它转成杯口向上，铃声就会震鸣. 因此，我们不妨仍然假定游戏者的触手探查到的乃是目前杯口统统向上的六边形. 这时，我们教游戏者把整个六角形一律转为杯口向下的状态. 结果所得的状态 S_5 标明在图 6 之中.

现在进入第二回合的第二次探查. 请记住 P_2 的探查办法是从每一只六边形中检查一只三角形. 如果游戏者触摸到的三角形中，有一只是非单纯的，那它必然就是例外三角形. 在此种情形，我们只要教游戏者把例外杯子转为杯口向上，这样做了以后，就把所有的状态都变成单纯，而这等于是说，此时的状态已经是一个完全立方体了. 既然它已是我们的目标，我们就可以告诉游戏者，第二回合到此结束，下面一步可以省略. 反之，情况可能是，游戏者所摸到的全是单纯状

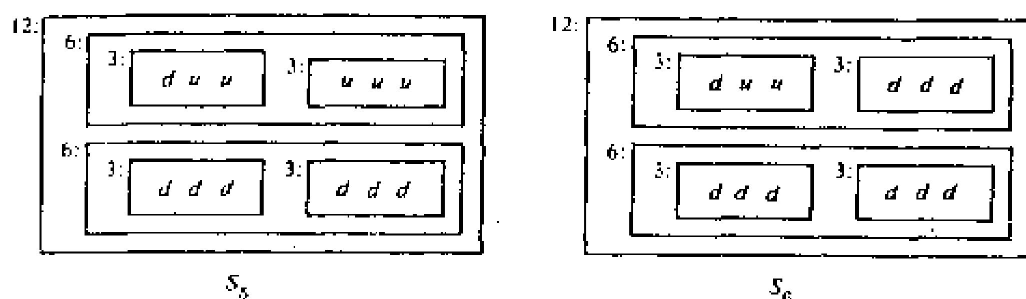


图 6

状态 S_5 与 S_6 的结构图.

态的三角形，它们中的一个单纯向上，另一个单纯向下；这时，我们教游戏者放过单纯向下的三角形不管，而把向上的三角形改为向下. 结果所得的状态 S_6，其结构图也在图 6 中给出.

我们现在进入第二回合的第三次，即最后一次探查，所用的模式是 P_3. 请回忆，P_3 应探查四只三角形中每只三角形的两个顶点. 目前，例外三角形的状态为〈duu〉，而所有其余的非例外三角形都是一致单纯向下. 请注意，游戏者不会搞不清楚究竟哪个三角形是哪个. 如果她触摸到的一个三角形中的两顶点都是向下的，那么该三角形必然是单调向下的，于是我们可以教游戏者对这些三角形放任不管. 在例外三角形中游戏者也有两只触手可用. 这两只触手触摸到的，要末是两只向上的杯子，要末一只向上，一只向下. 如属前者，我们把两只向上的杯子都转为杯口向下；如属后者，我们把向下的杯子转为向上. 所产生的净效果是保证那只例外三角形在下一状态 S_7 转为单纯状态. 然而 S_7 的所有其他三角形也全已是单纯的，事实上，它们都是一致向下的. 这样做了之后，游戏者已经迫使转台进入一种完全立方的状态，而这正是引理 3 的要求.

上面刚刚讲过的策略，我们将称之为上-下策略，这是因为在第一回合把除了例外杯子以外的所有杯子都转为杯口向上，而在第二回合则把除了例外三角形之外的所有三角形都转为向下之故. 当然，我们应对一切的 n 来证明引理 3，而不是只对 $n=12$ 加以证明. 其结果是，除了 n 是 2 的整数幂之外，我们都可以用上-下策略来对付. 唯有一件事情可能引起混乱，那就是记法. 反之，对 2 的整数幂来说，上-下策略无效，对于此种情形，我们将推迟到第 5 节再加讨论.

n 不是 2 的整数幂时引理 3 的证明　设 n 表示转台的大小，p 是 n 的最大素数因子. 由于 n 不是 2 的整数幂，所以 $p\geqslant 3$. 我们要为游戏者设计的取胜策略应具备以下性质：

要末它能响起铃声，要末迫使转台进入一个完全 p 次幂的状态 S，而这可以使用不多于 $(1-1/p)n$ 个触手来达到目的.

我们先来建立某些记法，设 n 具有 j 个素数因子，其中 p 是最大

的一个. 如果我们按递增次序书写 n 的素数因子，则有

$$n = p_1 p_2 \cdots p_j, j \geqslant 1, p_1 \leqslant p_2 \leqslant \cdots \leqslant p_j = p.$$

从 n 的因式分解出发，我们再来定义某些辅助记号 l_i 与 r_i; l_i 的值是若干个 n 的素数因子按递增顺序的连乘积，但第 i 个素因子不计在内，而 r_i 则是从第 i 个因子一直乘到第 j 个因子的连乘积. 对 $1 \leqslant i \leqslant j+1$，我们可以由下列式子正式地定义 l_i 与 r_i，即

$$l_i = \prod_{1 \leqslant k < i} P_k,$$

$$r_i = \prod_{i \leqslant k \leqslant j} P_k.$$

请注意，对一切 i，均有等式 $l_i r_i = n$ 成立.

为了理解上-下策略的一般情况而非特例，人们必须用一种相当复杂的方式来看待 n 边多边形. 如果从外面开始，我们可把 n 边形（等价的说法是 r_1 边形）看作 p_1 个相互穿插的 r_2 边形所组成，而每个 r_2 边形则由 p_2 个相互穿插的 r_3 边形所组成，如此等等，一般地说，每个 r_i 边形由 p_i 个互相穿插的 r_{i+1} 边形所组成. 最后，每个 r_j 边形由 p_j 个顶点所组成，也就是说，r_j 等于 p_j. p_j 处于这种层次结构的底层，它所起的作用犹似 $n=12$ 的场合中三角形所扮演的角色.

上面一节阐述了 r_i 的意义，我们把 n 边多边形分解为不同大小的多边形，其尺度即由 r_i 来表示. l_i 的作用则是告诉我们一共有多少个 r_i 边多边形. 特别地，对每个 i，n 边多边形正好确切地含有 l_i 个相异的 r_i 边多边形.

现在我们有可能对一般情况下的结构图有进一步的理解. 长方形的嵌套深度有 j 个层次，在每一层次，l_i 个相异的 r_i 边多边形，每一个都分解为 p_i 个相异的 r_{i+1} 边多边形. 图 7 试图表明，在一般情况下，第 i 层嵌套的结构图看上去将像些什么.

在一般情况下的上-下策略需要应用 j 个不同的探查模式，我们将称它们为 $P_1, P_2, \cdots, P_j$. 如同 $n=12$ 的情况一样，只要它们具有正确的结构图，我们可以不去顾问 P_i 究竟是什么具体的探查模式. 我们可以用图 7 来表明 P_i 所需要的第 i 层次嵌套的结构图. 特别地

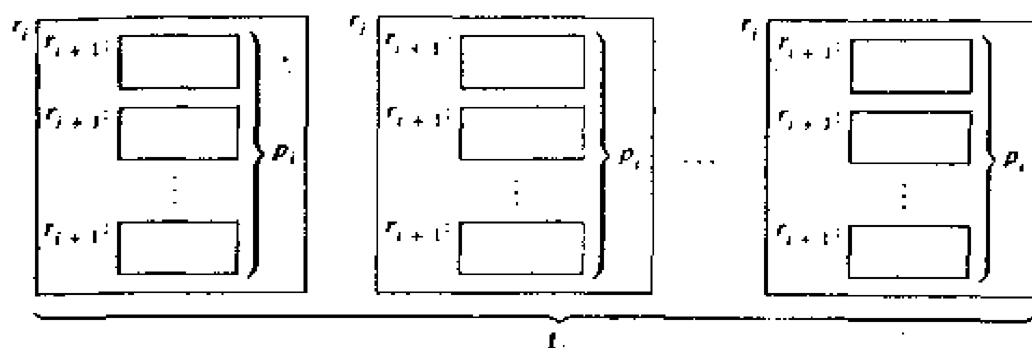


图 7

一般情况下第 i 层嵌套的结构图.

说,模式 P_i 应当探查组成每一个 r_i 边多边形的所有的 r_{i+1} 边多边形,除了一个之外. 为了方便起见,我们将把此种情况简单地说成:这种模式解决了 i 层嵌套问题.

我们现在任意选择模式 P_i,其条件是 P_i 必须能够解决 i 层次的嵌套. 下一步要加以验证的事是:游戏者必须拥有足够的触手来实现我们已选定的模式 P_i. 为了说明此事,请注意,解决 i 层嵌套的任一模式都恰恰用上

$$t_i(p_i-1)r_{i+1}=\left(1-\frac{1}{p_i}\right)n$$

个触手. 由于我们的游戏人准许使用 $\left(1-\frac{1}{p}\right)n$ 个触手(这里 $p=p_j \geqslant p_i$),她确实拥有足够的触手以便她使用模式 P_i.

我们现在建议游戏者执行第一回合,即依序地用模式 P_1 至 P_j 进行试探,把她所能触摸到的杯子都转成杯口向上. 模式 P_1 探查整个转台,即一个 r_1 边多边形除去一个 r_2 边形外;接着,模式 P_2 是要探查所有的 r_2 边多边形(除去一个 r_3 边形外). 这个过程将一直继续进行下去,即:用模式 P_i 把转台约仍可能含有一只向下杯子的那部分区域进一步限定到一个 r_{i+1} 边形. 于是,最后用 P_j 探查过之后,整个转台最多只可能具有一个杯口向下的杯子. 我们自然可以假定恰是只有一只向下的杯子,否则我们的游戏者已经赢了. 我们将使用一

个术语例外的，它既是指这只杯口向下的杯子，也是指包含它的 P_j 边多边形. 第一回合所起的作用是完成了上-下策略“上”的这部分要求：除了一只例外的杯子之外，它已使所有的杯口统统都向上了.

从我们选择模式 P_i，随之而来的是 P_i 要探查任意一个 r_{i+1} 边多边形（它的任意一些顶点要被探查）的所有顶点. 这意味着，对任意 $k>i$ 来说，P_i 事实上必须要末探查所有的 r_k 边形，要末一个都不查. 这一事实有助于说明我们所使用的术语. 为了只要检查一个 r_k 多边形的一部分：换言之，为了解决 k 层次嵌套问题，人们需要利用一个模式 P_i，这里的 $i\geqslant k$. 特别地，我们注意到，除了最后一个模式 P_j 之外，所有的模式 P_i 都必然要末探查所有的 r_j 边形（即每个 P_j 边形），要末一个都不查. 所观察到的现象也可以重新用短语描述为：对 $i<j$ 的模式 P_i 是完全的 p_j 次乘幂.

现在进入第二回合. 我们建议游戏者继续依次利用探查模式 $P_1,P_2\cdots,P_j$，但指示得更为复杂一些. 把转台看作由互相穿插的 p_j 边形所组成. 当我们开始作第二回合探查时，所有的 p_j 边形都已单纯向上，只有一个例外，它包含着一只口子向下的杯子. 在本回合的最后一个探查之前，我们将要去探查一切 p_j 边形而其中只有任意一部分可以被我们查到. 我们教导游戏者：除了最后一次探查之外，其他都按下列方针办事：如果你触摸到那个例外的 p_j 边形，把那个例外的杯子口子朝天，这样就使转台进到一个完全的 p_j 次幂，于是你已满足引理 3 的要求，第二回合的余下探查就可以省略；如果你触摸到的是一个单纯向上的 p_j 边形，这表明它不是那个例外的多边形，而且还不曾逆转过，那你就把这个 p_j 边形中所有的杯子都一律改为口子向下；如果你摸到一个单纯向下的多边形，这表明它已被逆转过，那就放任它不管.

我们能证明仅有一只 p_i 边形在第二回合的前 $j-1$ 次探查中得以逃避检查. 该项论证实质上与我们在上面用来分析第一回合的论证相同. 利用 P_1 所作的探查除去那些位于一个特殊的 r_2 边形中的 p_j 边形之外会碰上所有的 p_j 边形；而用 P_2 所作的探查则除了那些

位于一个特殊的 r_3 边形中的以外，将会碰上所有其余的；如此等等. 在利用 p_{j-1} 作过探查之后，除去那些位于一个特定的 r_j 边形($r_j=p_j$)中的以外，所有的 p_j 边形都将被探查到. 另外，我们可以假定，唯一例外的 p_j 边形没有被探查过，因为如果我们的游戏者有足够的运气探查了例外的 p_j 边形，她早已听从教导，矫正了杯子的朝向，从而获胜离开了. 因此，第二回合的前 $j-1$ 次探查已经触摸了所有的，非例外情况的 p_j 边形，并且都已把它们的杯子转为口子向下. 通过这样的动作，它们实现了上-下策略中“下”的状态.

现在可以讨论第二回合的最后一次探查了. 这时，转台含有一个例外的 p_j 边形，其中含有 p_j-1 只向上的杯子与一只向下的杯子，而所有其他的 p_j 边形都是单纯向下的. 在最后的探查中，我们要教导游戏者去利用模式 P_j，对每个 p_j 边形的 p_j-1 个顶点进行探查. 这里我们将抵达一个微妙之点. 迄今为止，我们并不需要利用 $p_j=p\geqslant3$ 的假设，但现在这个假设却是举足重轻的. 请注意，由于 $p_j\geqslant3$，游戏者必将从每一个 p_j 边形中摸到至少两只杯子，因此她必然会至少触摸到例外 p_j 边形中的一只向上杯子. 这将使她能够判断出，在 p_j-1 个触手中，哪一组是在触摸例外的 p_j 边形的. 我们在这里必须排除 $p_j=2$ 的情形，因为恰恰是在此种情况下论证将会失效. 如果 $p_j=2$，则唯一的 p_j 边形将是状态为$\langle du\rangle$的一条对角线. 如果游戏者探查这条对角线的一头，她有可能摸到杯口向下的一头，因此她就没有办法把这只口子向下的杯子同所有其他口子向下的杯子区别开来.

但在目前，由于我们已假定了 $p_j\geqslant3$，所以游戏者将有可能圈出例外的 p_j 边形. 我们教游戏者把所有的非例外 p_j 边形都放过不管. 对于例外的 p_j 边形，则有两种可能性. 要末她将会摸到那只孤零零的口子朝下的杯子，其时她只需把它颠倒一下，即可了事；要末她将会摸到 p_j-1 只口子向上的杯子；这时她可以把它们统统颠倒向下. 不论属于哪种情况，在下一状态时例外的 p_j 边形都将变成单纯状态，因而引理 3 的要求将得到满足. 这样也就完成了(对任一不等于 2 的整数幂的 n)引理 3 的证明.

第五节
上-翻策略

我们余下来的工作是要为蕴含在引理 3 中的、转台的大小为 2 的整数幂的情况提供一种取胜策略. 上文已经指出过，在此种情况下，上-下策略将无能为力. 因为在第二回合的最后一次探查中，游戏者将不可能定出那条例外的对角线. 在本节中，我们将为 2 的整数幂情形设计出一种名为*上-翻的策略*，它有点像是上-下策略，但却包含着一个技巧更为复杂的第二回合. 所谓"*翻转*"一只玻璃杯的意思是要把现在口子朝下的杯子转为朝上，现在朝上的转为朝下. 在上-翻策略的第二回合，我们将要教导游戏者，在每次探查中，把她所触摸到的每条非例外对角线统统翻转，而不是把它们转为口子向下. 反复进行的翻转将使转台进入一种奇妙的状态，各式各样朝上朝下的杯子排列成一种有趣的配置，而一个谨慎的游戏者将有可能从这样的结构中捞到好处. 为了给阐述上-翻策略的微妙性作些准备，我们需要研究圆圈字符行的某些事实.

对任意整数 k 与圆圈字符行 T，当且仅当 T 中的每一个记号与其右面第 k 个记号(点数时围绕圆周 T、周而复始)完全相同时，则 T 被说成是*以 k 为周期的*，或者说，k 是 T 的一个*周期*. 换言之，一个圈是 k 周期的，如果当它向右转过去 k 个位置后看起来完全相同的话. 显然，任意一个圈是 0 周期的；如果一个圈是 l 周期的，则它也必然是 $(-l)$ 周期的；如果一个圈既是 l 周期的，又是 m 周期的，则它也必定是 $(l+m)$ 周期的. 用抽象的语言来说，圈 T 的一切周期的集合构成一个整数的加法子群.

由于长度为 n 的圈恒为 n 周期的，故知每一个圈必有某些正整数周期. 我们将把一个圈的最小正整数周期称为它的*基本周期*，于是有下面的结果.

引理 4 任何圈的周期是其基本周期的倍数.

证明 基本周期的任一倍数肯定是周期.让我们来看逆命题,设k是一个圈的基本周期,m是某个其他周期,q是k与m的最大公约数.由于q可以记为某些整数a与b的下列等式

$$q = ak + bm,$$

因而q也必定是圈的一个周期.由于基本周期k是圈的最小正整数周期,所以必然是$q=k$.又因q能整除m,所以我们能断言m必然是基本周期的一个倍数.具有代数修养的读者会认识到这个证明其实就是整数构成一个主要理想域的证明.

周期的概念是相当标准的,实际上,它与完全幂概念有着紧密联系.请注意,如果一个长度为n的圈具有基本周期k,则k可整除n,于是这个圈将具有完全(n/k)次幂.但我们还是需要其他更吸引人的概念.这种另外的概念只是对恰有两种符号的圈才有意义,在目前的场合,这两种符号便是u与d.我们将把这两种符号称为:一个是另一个的对偶或补.于是,我们将把一个圈称为k重交替的,当且仅当圈中的每个符号都是其右边第k个符号之补.换言之,如把这个圈向右转过k个位置,其结果就好像它根本没有转,只是把每个符号换成了它的补.

如果一个圈是k重交替的,则它也必然以$2k$为周期,因为把某个东西求补两次等于是根本不求补.但是,也有可能圈的某个周期小于$2k$.例如,长度为六的圈(010101),这个圈是3重交替的,且其周期为6,但它同样也是以2为周期的.下面的引理将表明,如果k这个数是2的整数幂,那就不会有此种较小的周期.此种结果看来与我们要办的正事毫不相干,但请稍待片刻.

引理 5 如果一个圈是2^i重交替的,则其基本周期正好就是2^{i+1};换言之,一个2^i重交替的圈不存在较小的周期.

证明 2^{i+1}肯定是一个周期,于是圈的基本周期必能整除2^{i+1},所以它必是2的一个幂,但如果对某个$l \leqslant i$,2^l是一个周期,则2^i将也是圈的一个周期.由于没有一个圈能够同时为2^i周期与2^i重交替

的. 因而，基本周期必然恰恰就是 2^{j+1}.

把引理 5 放在一边，我们即可开始讲上-翻策略.

当 n 是 2 的整数幂时引理 3 的证明 设想我们的游戏者正在面对一只大小为 $n=2^j$ 的转台. 我们要为她设计一个策略，使她利用不多于 $n/2$ 只触手，即可使铃声震响或者迫使转桌进入一个完全二次幂的状态. 如果 $n=2$，只要进行一次浅显的试探即已足够. $n=4$ 的情况并不那样浅显，但是确实存在着一个四步策略可以解决问题. 请读者们参看马丁·加德纳在《科学美国人》1979 年三月号上的专栏文章里所介绍的原来的转台问题的解法中的前四步❶. 所以，我们只要集中考虑 $n\geqslant 8$ 的场合，或者说，$j\geqslant 3$ 的情形. 令人感到一些诧异的是，$n=4$ 的情形需要加以特别对待，然而这是必要的；因为我们所设计的上-翻策略恰恰在方桌子的情况下解决不了问题.

上-下策略与上-翻策略的第一个不同点在于 j 个探查模式 P_1，$P_2,\cdots,P_j$ 的定义. 就前者来说，我们满足于自由选择这些模式，只要 P_i 能解决 i 层嵌套问题就行. 而在后者，我们将选择具有这一性质的特殊序列，但也要具有一种简单结构. 说得更详细些，我们要精心安排模式 P_i，使之具有 2^{j-i} 组，每组具有 2^{i-1} 只连续的触手，相互之间被 2^{j-i} 组空白隔开，而每组具有 2^{i-1} 个连续的空白. 说得更正式些，我们可以由下列关系式

1. $$P_i=\langle(g^{2^{i-1}}h^{2^{i-1}})^{2^{j-i}}\rangle$$

来定义 P_i.

模式 P_1 由互相错开的 g 与 h 所构成，而模式 P_j 可以探查到一个半圆而留下其余部分未查. 现在考察 P_i 中每一组 h 的第 k 个触手(k 的范围为 $1\leqslant k\leqslant 2^{i-1}$). 这 2^{j-i} 个触手在一起可以探查构成一个 2^{j-i+1} 边多边形的两个 2^{j-i} 多边形中的一个，而另外的组成部分 2^{j-i} 边多边形则未被探查到，这是因为它正好处于 P_i 中每组 g 的第 k 个

---

❶ 译者注：鉴于《科学美国人》的过期杂志一般读者不易弄到. 有兴趣的读者可以参看译者所写的《神奇的转台》一文，该文发表于《少年科学》1990 年 1 月号.

空白位置.这就说明,由方程1定义的模式 P_i 确实能解决第 i 层次的嵌套问题.

上-翻策略的第一回合与上-下策略的第一回合是相同的;我们只要教游戏者去探查每一个模式 P_i,并把她所能触摸到的杯子统统使其杯口向上.与上-下策略中所作过的分析一样,我们得知,游戏者要末在这一回合的某一时刻赢得了整个游戏,要末在回合的结尾,转台变成只有一只向下的杯子,而其余 $(n-1)$ 只都是杯口向上.让我们把那个孤零零的口子朝下的杯子称为例外的杯子,而把包含它的对角线称为例外的对角线.由于这一论证中全部好戏都是在第二回合中出现的,我们将稍稍偏离一下通常的下标记法,使用符号 S_1 来表示第二回合开始时的转台状态.第一回合已迫使转台进入状态

$$S_1 = (du^{2^j-1}).$$

第二回合的首次探查同它惯于做的一样.我们教游戏者使用模式 P_1.如果她摸到一只朝下杯子,它必然就是那只例外的杯子.她把它转为朝上而立即赢了这场游戏.所以我们可假定她摸到的全是朝上的杯子,此时我们叫她把这些杯子悉数朝下放,这就迫使转台进入状态

2. $$S_2 = (dd(ud)^{2^{j-1}-1}).$$

在我们的游戏者继续深入第二回合时,我们教她找出一张纸片以便记录转台的状态.她的记录由方程2给出的状态 S_2 开始.有这张纸片在旁边,我们就能给游戏者交代一系列指示,它们将指导她进行第二、第三……次探查,但不包括第二回合的最后一次探查.设计这些指示的目的是要使某些条件的满足一直能够保持.设 S_i(这里有 $2 \leqslant i \leqslant j$)是做第二回合第 i 次探查时的当前状态,我们将维持下列条件:

3. 状态 S_i 与 2^{i-2} 交替的圈 T_i 仅仅有一处差异;

4. 在第二回合的第 $i-1$ 次探查刚刚做过之后,我们的游戏者可以把状态 S_i 记录在她的纸片上.

对 $i=2$,上述两个条件都能满足,因为状态 S_2 由方程2给出,

而且它与一个1-交替的圈只有一处差异(那只例外杯子置放之处).在第二回合的中间各次探查中我们的任务是要使这些条件继续得到保持.让我们考察第二回合的第i次(这里的i要满足$2\leqslant i<j$)探查,我们假定所有的条件迄今都能保持.对这次探查,我们教游戏者使用模式P_i.由于$i<j$,我们可像上-下策略一样推出P_i是一个完全平方;换言之,它得以探查任一根可探查其任一头的对角线的两头.我们首先要求游戏者区分她所触摸到的对角线.

根据条件3,她正在探查的转台状态S_i与2^{i-2}交替的圈T_i仅在那个例外位置有差异.圈T_i必定也是以2^{i-1}为周期的,且因$i<j$,它必定至少是一个完全四次幂.于是我们可断言,S_i的所有不是例外对角线的对角线必定是单纯的.因而我们的游戏者立即可以决定她是否正在探查例外对角线.如果她确实摸到它,我们教她使之变为单纯,其办法是把它的朝下杯子变为朝上,朝上杯子变为朝下.两种情况中究竟属于何者是无关紧要的.这样做了之后,就将使转台进入一个完全平方的状态,于是就满足了引理3的要求,第二回合余下的一些探查即可省略.

假定我们的游戏者并未探查到例外对角线,也就是说,她摸到的只是单纯的对角线.上-下策略教她把所有触摸到的单纯p_j边形统统转成向下,但在眼下这个策略中,我们教它把单纯向上对角线转为向下,而把单纯向下对角线转为向上.换言之,即把她所触摸到的每一条对角线统统翻转.这便是上-翻策略得名的来由.可是,它到底有什么好处?由于S_i除了例外杯子外是2^{i-2}交替的,所以除了例外杯子,它也是以2^{i-1}为周期的.把利用p_i所能摸到的每只杯子统统加以翻转,其效果实际上即是求补,并依然维持着长度为2^{i-1}的交替组不变.因此,翻转将保证,除了例外杯子之外,下一个状态S_{i+1}是2^{i-1}交替的.这就是说,翻转保证了在时刻$i+1$时,条件3仍然成立.

但条件4又将如何?请记住我们已假定游戏者并未探查到例外对角线,因而,她把触摸到的每一只杯子全部加以颠倒.既然状态S_i已写在她的纸片上,这就足以表明她可以毫不犹疑地了解到她正在

触摸的是 S_i 中的哪些杯子，也就是说，她的触手究竟摸在转台的哪处. 先考虑一下，她用模式 P_i 探查圈 T_i 时会发生什么情况. 由于圈 T_i 是 2^{i-2} 交替的，用每一组 2^{i-1} 个触手的试探结果将会得出 $X\bar{X}$ 的形状（这里 $X\bar{X}$ 是长度为 2^{i-2} 的字符行）. 现在，假定字符行 $X\bar{X}$ 当真在环绕 T_i 的两个不同地方出现（该两处由一个 l 位置的转动来分开），由于 T_i 也是以 2^{i-1} 为周期的，因此，整个圈 T_i 将会与旋转 l 步的圈重合. 换言之，l 必然是 T_i 的一个周期.

最后，我们能够看到引理 5 的关联. 从那个结果，我们推出 2^{i-1} 必然是 T_i 的真正基本周期，因而 l 必为 2^{i-1} 的一个倍数. 考虑 P_i 的结构，它只有两种不同的方式以使得圈 P_i 与 T_i 结合起来产生探查结果 $X\bar{X}$. 事实上，T_i 将恰好含有字符行 $X\bar{X}$ 的 2^{j-i+1} 份拷贝，而游戏者的触手必将要末摸到所有的奇数拷贝，要末摸到所有的偶数拷贝. 任何其他可能性都将使 T_i 具有一个非平凡的周期，而与引理 5 发生矛盾. 可是我们的游戏者实际上是在探查 S_i，而不是 T_i. 由于我们已经假设游戏者并未触摸到例外的杯子，这就消除了环绕 S_i 的她的触手的两种方式中的一种，剩下一种可能性了. 于是，我们的游戏者将能从第 i 次探查的结果中确切地得知她的触手究竟触摸在圈 S_i 的什么位置. 有了这个信息，她可以很容易算出圈 S_{i-1}，并把它写在纸片上.

我们业已证明，在整个第二回合的中间阶段，即上-翻策略的“翻”相中，我们的游戏者可以一直维持条件 3 与条件 4 不变. 现在我们要讨论第二回合的最后一次探查，此时她将要探查一个半圆. 首先，我们要证明游戏者有能力搞清楚她是否摸到例外的杯子. 这一次，她已不能通过察看非单纯的对角线来区别例外对角线，因为她只能探查每条对角线的一头. 但从适用于 $i=j$ 情况的条件 3，我们得知输入到第二回合第 j 次探查的转台状态 S_j 与一个 2^{j-2} 交替圈 T_j 仅有一处差异. 因而，我们的游戏者只要检查一下她的探查结果是否具有形式 $X\bar{X}$（X 为长 2^{j-2} 的字符行）即能肯定她是否触摸到例外的杯子. 说得更详细一些，如果她错过了例外的杯子，她的探查结果将

具有 $X\bar{X}$ 的形式；如果她触摸到例外的杯子，则她的探查结果将具有 $YdZ\ \bar{Y}d\bar{Z}$ 的形式，在这里，两个 d 中的一个即是例外的杯子.

设想我们的游戏者确是摸到了例外的杯子，她的下一步任务是要确定哪一个 d 是例外的，以便将其改正. 正是在这里，我们的假定 $j\geqslant3$ 起了决定性作用. 如果我们要对一只正方形转台使用上-翻策略，此时就会面临在 dd 组中决定哪一个 d 是例外杯子的问题，而这是没有办法判断的. 但我们已经通过一个特殊的论证来处理掉 $n=4$ 的问题，因此我们得以断言，两个字符行 Y 与 Z 中，至少有一个是非空集合，因此在例外杯子的当前两个候补者的紧邻，我们的游戏者确有可能触摸. 请注意紧邻将是互补的. 由条件 4，我们知道，游戏者的纸片上已经写着确切的圈 S_j 的状态. 根据这个圈，容易推出例外杯子的左、右紧邻到底处于什么状态. 这种观察将使我们的游戏者得以决定哪一个候补者是真正的例外杯子. 我们教她翻转那只杯子. 这就迫使例外对角线进入单调状态 $\langle uu\rangle$，从而满足了引理 3 的要求.

仅有一种情形需要处理了. 我们现在假定，第二回合的最后一次探查中，我们的游戏者并未摸到例外杯子，也就是说，她的探查结果可概括为形式 $X\bar{X}$. 此种情形下，仅在例外位置与状态 S_j 有所差异的圈 T_j 必然由公式 $T_j=\langle X\bar{X}X\bar{X}\rangle$ 给出，再次从引理 5 加以论证，我们能够证明字符行 $X\bar{X}$ 仅在两个明显位置作为 T_j 的子行出现. 因此，它作为 S_j 的子字符行只能出现一次. 这意味着我们的游戏者将能唯一地确定她的触手围绕 S_j 的位置. 由于她正在触摸一个半圆周，而她并未触摸到例外的杯子，所以她一定是在触摸例外对角线的另一个顶点. 我们教游戏者看一看她的纸片，确定哪一只她所摸到的杯口向上的杯子是与那只例外杯子同处于一根对角线的两头的. 只要把这只朝上杯子颠倒一下，即可迫使例外对角线进入单纯状态 $\langle dd\rangle$，从而满足了引理 3 的要求.

我们业已完全决定了函数 $f(n)$ 的值，但这并不意味着我们已可把转台问题搁置一边. 反之，我们在引言中已提到过，本问题的其他几种推广形式是值得探讨的，即便我们已讨论过的这种形式也值得

进一步研究. 例如,对具有 k 只触手(k 等于或大于 $f(n)$)的游戏者,怎样设计较短的取胜策略也许会引起人们的兴趣.

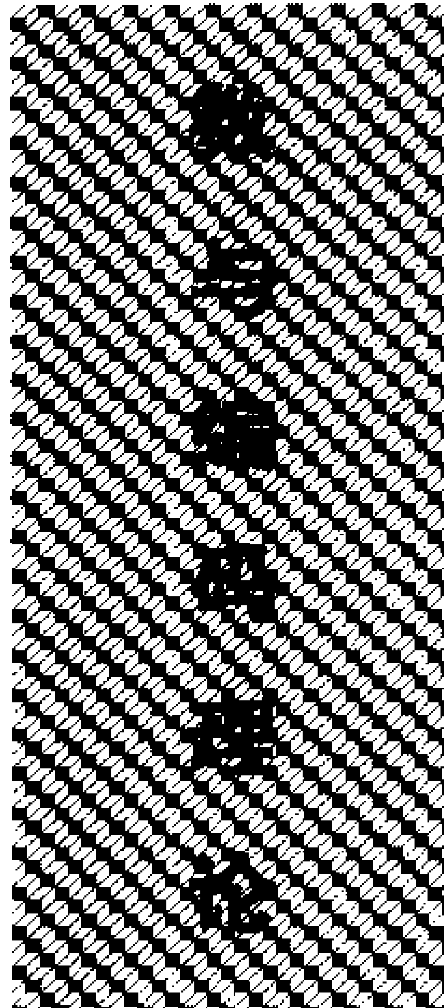
我们讨论过的策略设计技巧将能为我们提供一族策略的长度的事前估量. 若令 $n=a_1a_2\cdots a_j$ 是 n 的一个因式分解,其中每个因子 a_i 都满足条件 $a_i\geqslant 2$;假定为了方便起见,所有的因子 2(如果有的话)都是排在前面. 如果 $a_i>2$ 时使用上-下策略,$a_j=2$ 时使用上-翻策略,我们得以将问题的规模为 n 的策略设计降低为问题规模为 n/a_i 的策略设计,而使用的探查次数最多为 $2j$ 次,使用的触手数不超过

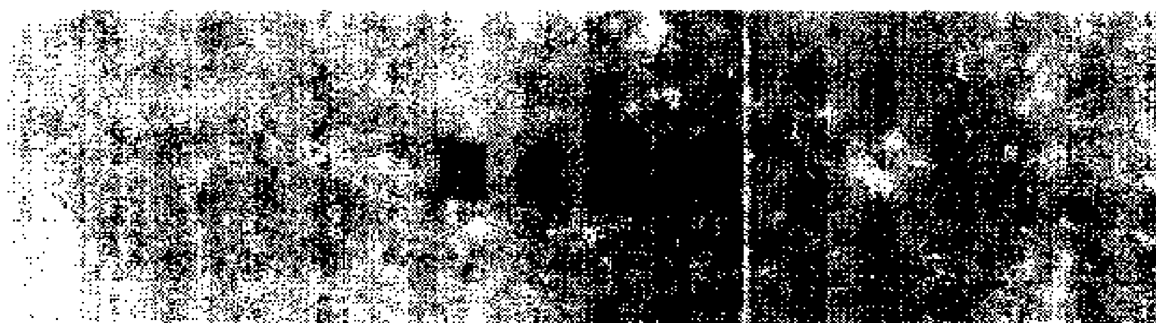
$$K=\max_{1\leqslant i\leqslant j}\left(1-\frac{1}{a_i}\right)n.$$

于是,我们可以为大小是 n 的转台设计一套取胜策略,使用的只是 K 只触手,探查次数至多是 j^2+j 次. 这样也就使人们可以权衡得失:或者是把 a_i 取得大些,这样可以使 j 较小,并使得取胜策略较为短些,但需要较多的触手;或者是把 a_i 取得小些以使 j 较大,这样取胜策略将会冗长一些,但是触手数将可用得少些. 另外,把这些策略稍为缩短一些并非难事,譬如说,从上-下策略的一步到下一步,时常可以捎带信息而免除探查. 但是我们还是要问:在最不利情况下,除了我们已讲过的办法之外,是否还存在其他策略,以便用少得多的探查来实现目的?

添加在证明中的注记:迪特·刘易士(Ted Lewis)与斯蒂芬·惠拉德(Stephen Willard)独立地决定了函数 $f(n)$. 他们的文章已发表于《数学杂志》(Mathematics Magazine)1980 年 5 月号(53 卷第 3 期,第 174—179 页),题为《转台》. 当 n 不等于 2 的整数幂时,他们使用了一种经过改进的上-下策略,把信息从上-下态的一步捎带到下一步,这可使游戏者得以利用大体上一半左右的探查次数来实现目标. 另外,他们发现了一种与上-下策略非常接近的策略来对付 n 是 2 的整数幂的情形:第一回合中把所有杯子转为口朝上,第二回合的绝大多数场合使其转为向下. 翻转手段仅限于在第二回合的倒数第二个探

查中用一用，它只是把一条对角线翻转了一下.





● 斯坦福大学

□ 唐纳德·E·克努特(Donald E. Knuth)

利奥波德·克隆内克(Leopold Kronecker)说道[10]:"上帝创造了整数,其他东西是人的产物". 如果克隆内克说得果真不错,那么由于拥有神奇力量而把非整数说成是超自然数,就将是异端邪说. 另一方面,数学家们通常把非负整数{0,1,2,…}称作自然数,因此,任何超自然数也必然是自然数. 本文目的是想讨论"超"自然数的表示法,所谓"超",乃是指它们是极其庞大的数,如果采用传统记法,就得使用许多上标才能表达.

要超越已知宇宙的大小,并不需要极其庞大的数. 譬如说,我们考虑的是一个每边长度为 400 亿光年的立方体,而在其中装入极其微小的、10^{-13} 厘米×10^{-13} 厘米×10^{-13} 厘米的立方体(每个微小立方体远远小于一个质子或一个中子),则小立方体的总数将小于 10^{125}. 这个数量说不上是超人的,它只是 125 位而已.

伟大的阿基米德似乎是讨论极大数存在的第一个人,他的著名论文《沙粒计数》[1]得出结论,不到 10^{63} 颗沙粒便足以填满他那个时代所定义的宇宙. 此外,他还进一步介绍了一个记数法,通过这种办法,甚至可以谈论大到 $10^{80,000,000,000,000,000}$ 的数目.

英语里头根本没有任何名词来表达如此庞大的数目. 我们怎样为大数的表达提供一种系统记法,考虑这样的问题无疑是很有教益

的.在目前这个通货膨胀时期,我们也许不久又需要编造新的单词来表示货物的价格;例如,在1923年时,德国发行的单张邮票,其面值竟达到500亿马克,然而这种邮票几乎是一钱不值的.

当我们停下来检查数的传统命名时,事情一下子就很清楚:这些命名都是"人为的产物",它们本应设计得更完善一些,譬如说,完全废弃"千"这种单位,"百"的下一个单位是"万"(10^4),也许会更好.因为,像1984这样的数,传统读法是"十九百八十四",而不是"一千九百八十四"[1]."万"的使用将为我们提供一批相当满意的叫法,直到10^8为止.例如,素数9999,9989即可读作"九十九百九十九万九十九百八十九".

我们所需要的,万的下一个单位是10^8.让我们大家一致同意把它称作一"元"(发音为mile-yun),类似的英语单词当然也可以用,例如"百万富翁"这个词.一元的下一个单位称作"二元",一个二元相当于10^{16}.使用这个并无歧义的名词"二元"显然是一大进步,因为人们根本弄不清楚目前正在使用的"billion"这个词究竟代表怎样一个数(英国人与德国人认为它相当于一百万个一百万,而美国人与法国人则认为它是一千个一百万,即十亿).这个新名词还有另一个优点,因为我们的国债总额只是此数的一个极小分数,也可使我们略为宽慰一些.我们的记数法将按如下方式进行下去:

10^{32}	三元	10^{2048}	九元	10^{131072}	十五元
10^{64}	四元	10^{4096}	十元	10^{262144}	十六元
10^{128}	五元	10^{8192}	十一元	10^{524288}	十七元
10^{256}	六元	10^{16384}	十二元	$10^{1048576}$	十八元
10^{512}	七元	10^{32768}	十三元	$10^{2097152}$	十九元
10^{1024}	八元	10^{65536}	十四元	$10^{4194304}$	二十元

[1] 原注:我们也可以考虑把"十九"改称"一十九",等等;但这也许会使我们的记法过分逻辑化了.(逻辑味道太重了.)

例如，一副扑克牌的洗牌法总数为 52！＝8065∷8175,1709;4387,8571：6606,3685;6403,7669;;7528,9505;4408,8327：7824,0000;0000,0000，我们把这个数按四位，八位，十六位分组，并使用不同的标点符号. 如果照我们在上面所建议使用的记数法，则这一数目的英语名称将是“八十百六十五个“四元”八十一百七十五万十七百九“一元”四十三百八十七万八十五百七十一个“二元”六十六百六万三十六百八十五“一元”六十四百三万七十六百六十九个“三元”七十五百二十八万九十五百五“一元”四十四百八万八十三百二十七个“二元”七十八百二十四万个“一元”. ”在我们目前使用的美国语言中，这个数目是没有办法表述的，除非我们依赖如下的说法：八十千六百五十八个“vigintillion”一百七十五个“nonillion”……八百二十四个 trillion，因为在普通辞典中，对于超过一个 vigintillion（相当于 10^{63}）的大数，除掉 centillion（相当于 10^{303}）之外，并没有一一为它们提供名称.

如果在我们的语言中对上帝的一切创造物都能表述的话，我们当然需要超越二十元，甚至一百元都要超过的. 如果我们把上述模式加以外推，并采用拉丁式记法的话，那么下一个单位就应是二十一元. 于是，由下式

$$\begin{aligned}&80,000,000,000,000,000\\ &= 2^{56} + 2^{52} + 2^{51} + 2^{50} + 2^{45}\\ &\quad + 2^{44} + 2^{42} + 2^{41} + 2^{40} + 2^{39}\\ &\quad - 2^{36} + 2^{33} + 2^{32} + 2^{30} + 2^{29}\\ &\quad + 2^{28} + 2^{27} + 2^{26} + 2^{25} + 2^{19},\end{aligned}$$

真正懂得我们所建议的记数法的读者将能读出阿基米德书中的最大数目 $10^{80,000,000,000,000,000}$（本文末尾有其答案）.

对给出新单位来说，我们迟早会把所有的拉丁名字统统用光，因为罗马人从来把数字算得很大. 即使罗马学者也采用了阿基米德的记数方案，在对下面之类的单位

$$10^{2^{10^{80,000,000,000,000,003}}}$$

进行命名时，也仍会遇到麻烦. 不过，我们可以避开这个问题，只要对每个基本单位 $10^{2^{n+2}}$ 按下列办法命名就行.

对很大的数 n，我们把它唤做“latin(n 的拉丁名称，空位 0 不算) yllion”. 例如，$10^{2^{100000000000000002}}$ 即可叫做“latinbyllionyllion”. 通过这种方式，我们即可叫出所有自然数的英文名称，而不管它们是如何庞大. 例如，有个自然数的名称为“latinlatinlatinbyllionyllionyllionyllion”，读者们能推算出它的大小吗？（参阅本文末尾的答案.）

讲到这里，读者们也许会问：“究竟为点啥？通常的十进制或二进制记数法已经为我们提供了一个既便于表达又适合计数的办法，有谁会去计较庞大数目的名称呢？”不错，数的名称对我们大脑皮层掌管语言功能的那部分要比执掌计算功能的那些灰色细胞更为重要. 只是由于阿基米德这样的伟大人物曾经写过一些文章，这一点本身并不能证明它具有什么科学价值. 可是我们的讨论不会全无价值，因为它使得我们对讨论一个更为重要的问题有了准备，而在解决该问题时需要用上类似的概念.

在本文的余下部分，我们所要讨论的问题是：怎样通过 0 与 1 的序列来表示任意大的自然数，并使之满足以下两个条件：

1. 每一个数的表示决不是其他任何数的表示的“前缀”（或“词头”）.

换言之，如果某个数的表示为序列“01101”，则别的数目决不能有以“01101”开始的表示. 数的表示尚需满足的第二个条件是

2. 每个数的表示应比其他任何较大的数的表示具有较小的“辞典长度”.

所谓较小的辞典长度意味着它应该在辞典中较早出现；例如，01101 的辞典长度要比以“1”打头的序列为小，它的长度也小于以“0111”，“011010”，以及“011011”打头的序列.

这一问题的重要性在于：我们经常需要通过一串数字同计算机打交道. 例如，这串数字是一些指令，它告诉计算机要做些什么. 关键问题是计算机必需懂得一个数在何处结束，下一个数在何处开始. 条

件 1 能保证做到这一点，因为计算机可以从左至右进行读数，不至于混淆不清. 条件 2 虽然并不具有如此的关键性，但它有一项相当优美的性质. 它蕴含着表示法的保序性，就是说，如果一串数字比另一串数字的辞典长度小，则第一串数字的表示比起第二串数字的表示来，辞典长度也来得小.

如果我们只要表示很少几个数目，那么条件 1、2 是很容易得到满足的. 例如，假定我们事先知道要用到的仅仅是自 0 至 7 几个数目，那就可以使用同样的字长来表示它们，即：000，001，010，011，100，101，110，111. 不过，我们需要的是对一切自然数均能进行编码，而不是为数很少的几个小数目.

人们想到的第一种正确解决方案也许是一种“一元”表示法，即把 0 表示为“0”，1 表示为“10”，2 表示为“110”……以此类推，n 可表示为“1^n0”，也就是 n 个 1，后面再跟上一个零. 零的作用犹似计算机必需读入的相继数目中间的休止符或“逗号”. 此种表示法显然能够满足条件 1 与 2，但它不大可能是一种实际可行的办法，即便对中等大小的数也不行，因为它将需要用 $n+1$ 位才能表示 n. 我们真正需要的是一种能满足条件 1 与 2，而又尽量紧凑的表示方法.

第二种最简单的解决方案也许是基于常见的、整数的二进制记法，即{0，1，10，11，100，101，…}. 这时，无论条件 1 或条件 2 均不能满足，但我们可以把此种表示法加以改进，其办法是：在数的二进表示法的前面加上一个表明字长的数字序列，例如：

0→00	8→1110000
1→01	9→1110001
2→100	10→1110010
3→101	15→1110111
4→11000	16→111100000
5→11001	31→111101111
6→11010	32→11111000000
7→11011	63→11111011111

64→1111110000000

127→1111110111111

128→111111100000000

255→111111101111111

256→11111111000000000

511→11111111011111111

512→1111111110000000000

1000→1111111110111101000

一般地，对 $n \geqslant 2$，如果数 n 的二进表示为 1α（α 是 0 与 1 的序列），则 n 的新表示法便是

$$1^{|\alpha|}0\alpha,$$

这里的 $|\alpha|$ 表示 α 的字长[1]. 如果用二进制记法写出 n 时需要用 m 位，则 n 的新表示法就将需要 $2m-1$ 位. 我们粗略地估计，表示数 n 所需的位数大致需要加倍，才能毫无混淆地表明每一个表示要在何处结束.

这一解答方案仍可加以改进，但在我们对之继续探讨以前，值得指出一下，我们的问题实质上相当于一种**猜数游戏**：一人任意认定一个自然数，另一人试图猜出它. 该游戏的规则有点像所谓的"二十个问题"：对非负整数 n，猜数者只能发问"你的数是否小于 n?"而另一人只能回答"是"或"否". 当然此种游戏可以比 20 个问题持续更久，因为根据 20 个是或非问题的回答将无法区别比 2^{20} 更多的数，可是准许认定的秘密数目却有无穷多个. 虽然如此，人们的一般想法是要尽可能快速地把数目猜出来.

此种猜数游戏与序列表示之间的联系是不难看出的. 猜数者能猜出一切自然数的任何办法都将是满足条件 1 与 2 的某种解决方

---

[1] （译者注）：让我们举一个实例说明一下. 数 31 的二进表示为 11111，由此可见 α 相当于 1111，其字长为 4，于是在写出 31 的新表示法时，须将 1 重复 4 次，即得 111101111. 其他数目可依此类推.

案：我们只要简单地把 n 的表示看作为一串“是或否”问题的回答，设想秘密认定的数目是 n，0 代表“是”，1 代表“否”．反过来，只要给出满足条件 1 与 2 的解决方案，我们即可构造出一个对应的猜数策略；只要我们允许猜数者提出一个笨拙的问题“你认定的那个数小于无穷大吗？”（其回答当然永远为“是”；这种情况意味着，在满足条件 1 与 2 的某些解决方案中，每种表示都是以 0 开头的）．

我们在上面最早提到的一元方案相当于猜数者采用以下办法提出问题：

你的那个数小于 1 吗？

你的那个数小于 2 吗？

你的那个数小于 3 吗？

……以此类推，直到他碰到第一个“是”的答复为止．我们的第二种方案要聪明得多，它以下面的一串问题开始：

你的那个数小于 2 吗？

你的那个数小于 4 吗？

你的那个数小于 8 吗？

这样以此类推……直到第一个“是”的答复暴露了那个秘密数的数量级，接下来就可用“二分法搜索”把那个数揪出来．

基于猜数游戏与表示问题的此种等价关系，即可明了，猜数者的一种良好策略即对应于数的一种紧凑表示．因此我们去探索一种优秀的表示法，实质上就无异于去探索一种良好的猜数技巧．

顺便说一句，猜数游戏并不是浅薄无聊的，它有重要的实际应用．例如，给出一个多项式 $f(x)$，如当 $f(0)$ 为负值，又对足够大的 x，$f(x)$ 为正值时，确知 $f(x)$ 恰有一个正根，则当且仅当 $f(n)$ 为正时，此根才是小于 n 的．因此，猜数的好办法势将得出一种不必计算 f 的导数，而能确定根的位置的有效算法．

如果我们能够意识到这实质上是利用一元办法来表示 n 的二进表示法的长度，那么上面所说的第二种解决方案还可改进，因为二进的办法还可用其他办法来取代！换言之，由于对 α 的长度进行编码

的序列 $1^{|\alpha|}0$ 仍有可能改用更紧凑的 $|\alpha|$ 的表示来取代. 反复运用以上概念,我们将能得到庞大数字的、越来越简短的表示法,而最终将导致如下的递归方法:先猜一下数 n 的二进制位数,再用二分搜索法来确定 n 的准确数值. 猜测 n 的二进制位数时,仍可递归地应用同样办法,于是我们可以先猜 n 的位数里头的位数,再猜 n 的位数中的位数中的位数,……如此等等. 由此可见,递归方法中的前几个提问是:

你那个数小于 1 吗?

你那个数小于 2 吗?

你那个数小于 4 吗?

你那个数小于 16 吗?

你那个数小于 65536 吗?

……如此继续进行下去,直至答复为"是". (请注意 $2=2^1$, $4=2^2$, $16=2^4$, $65536=2^{16}$, 下面一个问题将涉及 2^{65536}.) 按此种方式应用递归方法,直至找到了那个秘密认定数的上界,然后再一层层地解开递归. 如果秘密数确实非常巨大,猜数者将能比其他方法更快地猜到它.

可用极其简单的方式来定义与递归猜数法相对应的递归表示法:令 $R(n)$ 为表示 n 的 0 与 1 的数字序列. 则

$$R(0) = 0;$$

$$R((1\alpha)_2) = 1R(|\alpha|)\alpha,$$

这里的 $(1\alpha)_2$ 是二进制表示为 1α 的数,而 $|\alpha|$ 是序列 α 的长度. 一些较小的整数可表示如下:

0→0	7→1110011
1→10	8→11101000
2→1100	9→11101001
3→1101	10→11101010
4→1110000	15→11101111
5→1110001	16→111100000000
6→1110010	31→111100001111

32→1111000100000

63→1111000111111

64→11110010000000

127→111100101111111

128→111100110000000

255→111100111111111

256→11110100000000000

511→11110100011111111

512→111101001000000000

1000→111101001111101000

容易看出,这种表示法满足条件 1 与 2. 怎样把递归方法转变为迭代方法,对此感兴趣的计算机科学家将乐于寻找一种简单的非递归算法,使一旦给出 n 的表示后能立即得出其数值. 此问题的答案请参阅本文的末尾.

一个大数 n 的递归表示大致相当于此数的二进表示之半. 例如,在二进制方案下,数 $2^{65536}-1$ 的表示为序列

$$1^{65535}01^{65535}$$

其长度是 131071,而在递归方案下,其表示

$$1^{5}01^{65554}$$

仅有 65560 位长度. 反之,当 n 很小时,要清算数值,则递归方案着实需要一些功夫,在 2 与 127 之间的各数(包括 2 与 127 两数本身在内),二进表示要比递归表示更短. (仅当 $n=0$ 或 $n\geqslant512$ 时递归办法才优于二进办法.)因此,在很多应用中,宁愿使用二进方案.

倘若我们注意到除了 0 以外,使用递归法的一切序列都以 1 开始,则我们就有可能改进较小数目 n 的递归表示法. 如果我们只需表示严格正数,则不妨省略掉数列开头的那个"1". 另外,自然数与严格正数是一一对应的,故必存在一种表示 $Q(n)$ 以使得

$$1Q(n)=R(n+1).$$

应用此种经过修正的递归方案,我们有:

$$0 \to 0$$
$$1 \to 100$$
$$2 \to 101$$
$$3 \to 110000$$
$$4 \to 110001$$

以此类推;仅当 $n=1,3,4,5,6,7$ 以及 $15\leqslant n\leqslant 63$ 时,二进方案才优于本方案.同样的变换办法也可以应用于 Q 方案,按照同样的理由,编造出一种 P 方案,以使得

$$1P(n) = Q(n+1),$$

于是就有

$$0 \to 00$$
$$1 \to 01$$
$$2 \to 10000$$
$$3 \to 10001$$
$$4 \to 10010$$

如此等等;这时,仅当 $n=2,3,6,7,62,63$ 以及 $14\leqslant n\leqslant 31$ 时,二进方案才显得略为优越.

按照某种意义来说,如有可能,我们当然希望能找到最优策略,但是很难确切说定这种提法究竟意味着什么.二元策略看起来要比一元策略好得多,但即使是一元策略,它在 $n=0$ 时也能胜过二元策略.下面我们将要看到一种不可避免的情况.没有一种猜数游戏的策略能胜过别的——所谓"比别的策略好"其意义是指,对所有的 n 来说,为了确定 n,前者要提问的问题要比后者提的少,当然要剔除一些笨拙的问题,即可从前面的问题的答案中容易推出其"是"或"非"的.如果对某些 n 来说,一种已排除笨拙问题的策略要优于另一种策略,则对另外的一些 n,后者又会优于前者.总之,你是有赢有输.

为了分析这些方案究竟好到何种程度,我们有必要进一步定量化(请读者们注意:本文余下部分所包含的内容虽属初等,但有时使用的技巧相当精妙).如果 $n\geqslant 1$,我们把满足不等式

$$2^{\lambda n} \leqslant n < 2^{1+\lambda n}$$

的唯一自然数记为λn. 故若在二进制记法中 $n=(1\alpha)_2$,则数λn 是$|\alpha|$,即 α 的长度. 为了方便,可定义 $\lambda 0=0$,因此只要 n 是自然数,λn 也是自然数. 我们又可将 $\lambda(\lambda n)$记作 $\lambda\lambda n$,如此等等. 进而 $\lambda^m n$ 代表 λ 函数的 m 重反复,即 $\lambda^0 n=n$, $\lambda^3 n=\lambda\lambda\lambda n$, $\cdots$. 最后,我们也可用 $\lambda^* n$ 以表示使 $\lambda^m n=0$ 的最小整数 m.

可以很容易地应用此种函数来代表 n 的表示的长度. 令 $c(n)$为此种长度,即表示 n 的"成本"(或代价). 在猜数游戏中,$c(n)$便是猜出一个秘密认定数所需要的问题个数. 一元猜法有着很大的成本

$$c_1(n) = n + 1,$$

而二进制策略则把它降低到

$$c_B(n) = \begin{cases} 2 & \text{如 } n = 0 \text{ 或 } n = 1; \\ 2\lambda n + 1 & \text{如 } n > 1. \end{cases}$$

递归策略的成本为

$$c_R(n) = \lambda n + \lambda\lambda n + \lambda\lambda\lambda n + \cdots + \lambda^* n + 1,$$

这里,由"…"所代表的无穷级数实际上是有限的,这是由于当 $m \geqslant \lambda^* n$ 时,$\lambda^m n=0$.

最后,经过修正的递归策略之成本为:

$$c_Q(n) = c_R(n + 1) - 1,$$

$$c_P(n) = c_R(n + 2) - 2.$$

这些公式验证了我们以前曾经说过的话:当 n 很大时,递归策略的成本大约是二进方案的一半.

如果对应的猜数游戏从不提问愚蠢的问题(答案已知或答案可从已知事实中推出的问题),则此种表示办法称作是*无冗余的*. 其意思是:如果 α 是 0 与 1 的任意序列,而 α 不是任何 n 的表示,然而 α 却又是一个整数的表示的词头,则 $\alpha 0$ 与 $\alpha 1$ 都将作为表示的词头出现. 一切合理方案都是无冗余的,因若 $\alpha 0$ 作为词头出现而 $\alpha 1$ 则不然,那我们只要删去 α 后面的 0(当 α 作为词头出现时),即可既不违反条件 1 或 2,又能缩短某些整数的表示.

无冗余表示的成本函数满足一个重要的算术等式:

事实 1 设 $c(n)$ 为一种无冗余表示法中的成本,则必有

$$\frac{1}{2^{c(0)}}+\frac{1}{2^{c(1)}}+\frac{1}{2^{c(2)}}+\cdots=1.$$

[证] 事实上,若 α 是 n 的表示,我们有等式

$$\frac{1}{2^{c(0)}}+\frac{1}{2^{c(1)}}+\cdots+\frac{1}{2^{c(n-1)}}=(.\alpha)_2$$

(对二进制中的一切 n 均成立). 显然当 $n=0$ 时上式仍是对的. 因为在 0 的无冗余表示中不可能有 1. 设 β 是 $n+1$ 的表示,我们需要证明

$$(.\alpha)_2+\frac{1}{2^{|\alpha|}}=(.\beta)_2.$$

由只使用 1 的序列来表示的数是没有的,因为在辞典顺序上接在序列 1^m 后面的只是那种作为词头的序列. 于是,$(.\alpha)_2+2^{-|\alpha|}$ 小于 1,而事实上,$(.\alpha)_2+2^{-|\alpha|}=(.a_1a_2a_3\cdots)_2$是大于$(.\alpha)_2$又没有把 α 作为词头的二进制数中最小的一个数. 如果$(.\beta)_2=(.b_1b_2b_3\cdots)_2$ 不等于$(.a_1a_2a_3\cdots)_2$,令 j 是能使 $b_j\neq a_j$ 的最小的 j,而 k 是使 $a_k=1$ 的最大的 k,我们有$(.b_1b_2b_3\cdots)_2>(.a_1a_2a_3\cdots)_2$,因而 $b_j=1,a_j=0$. 如果 $k<j$,则序列 $b_1\cdots b_{j-1}$ 是一个冗余的词头,因为 $b_1\cdots b_{j-1}1$ 是作为 $n+1$ 的表示的词头出现,然而 $b_1\cdots\cdots b_{j-1}0$ 从来不作为词头出现. 如果 $k>j$,则序列 $a_1\cdots a_{k-1}$ 是一个冗余的词头,这是因为 $a_1\cdots a_{k-1}0$ 作为数 n 的表示的词头出现,但 $a_1\cdots a_{k-1}1$ 从来不作为词头出现.(请注意 k 是在 α 中最右面的 0 的位置.)

容易验证,对一切 k 来说,1^k 都作为词头而出现,如果现在 1^k 是 n 的表示的一个词头,则和式

$$2^{-c(0)}+2^{-c(1)}+\cdots+2^{-c(n-1)}$$

将介于$(.1^k)_2=1-2^{-k}$ 与 1 之间,因而当 $n\to\infty$时,无限项和式将收敛于 1. 证明完毕.

(此证法同时利用了 1 与 2 两条性质,如果只假定性质 1,则结果就将不真. 例如,下列步骤将对非冗余方式中的每个整数 n,以及

任意固定的 $k\geqslant 2$，产生一个长度为 $n+k$ 的表示：先用 0^k 表示 0，然后对 $n=1,2,3,\cdots$ 寻找一个序列 α，以使得(a)α 曾作为某些小于 n 的数的表示的词头而出现，但并没有作为数的表示而出现；(b)$\alpha 0$ 与 $\alpha 1$ 都不曾作为词头出现；(c)α 是满足(a)与(b)，并尽可能最短的序列. 于是任意长度为 $n+k$，有着以前从未出现过的 $\alpha 0$ 或 $\alpha 1$ 作为词头的序列都可以作为 n 的表示. 例如，$k=2$ 的这样一种表示法将由下面一些数开始：

0→00	3→11000	6→11100000
1→100	4→011000	7→010100000
2→0100	5→1010000	8→0111000000.

另一方面，对一切满足性质 1 的表示，不难证明下列不等式

$$\frac{1}{2^{c(0)}}+\frac{1}{2^{c(1)}}+\frac{1}{2^{c(2)}}+\cdots\leqslant 1.$$

这一关系，通常叫做克雷夫特(Kraft)氏不等式〔9〕)

也存在着事实 1 的逆命题，即

事实 2 设 $c(0),c(1),c(2),\cdots$是正整数的非减序列，且有

$$\frac{1}{2^{c(0)}}+\frac{1}{2^{c(1)}}+\frac{1}{2^{c(2)}}+\cdots=1,$$

则必存在着一种具有成本函数 $c(n)$的非冗余表示法.

［证］ 如果 n 的表示 α 可通过式子

$$2^{-c(0)}+2^{c(1)}+\cdots+2^{-c(n-1)}=(.\alpha)_2$$

定义，且 $|\alpha|=c(n)$. 我们即可得到具有所需性质的方案. 证明完毕.

若 X 是一种无冗余的表示方案，其成本函数 c_X 是非单调的，则我们可以重排这些整数以得到具有同一成本，但按非减顺序整理的其他表示办法. 事实 1,2 现在表明确实存在着一种拥有这些整理过的成本的无冗余表示. 拥有单调成本函数的表示方案可称标准方案，这是由于绝大多数应用都宁愿要关系式 $c_X(n)\leqslant c_X(n+1)$(对一切 n 而言).

让我们设法寻找可能有的最佳方案(重点放在渐近问题上,也就是放在极大整数的有效表示上)以结束我们的研究.很明显,二进方法比一元方法更为有效.我们也同样认为递归方法优于二进方法,因为在表达大数方面,前者比后者有更为卓越的性能.以上这些实例向我们暗示采用以下的定义:"若对一切大数 n,$c_X(n)\leqslant c_Y(n)$,而对无限多个 n,$c_X(n)<c_Y(n)$,则具有成本函数 $c_X(n)$ 的表示方案 X 优于另一个具有成本函数 $c_Y(n)$ 的表示方案 Y".根据这一定义,若 X 优于 Y,而 Y 优于 Z,则 X 优于 Z.

显然递归方法 P,Q,R 都优于二进方法 B,而 B 优于一元方法 U.然而,当我们试图比较三种递归方法的优劣时,事实表明没有一种方法比别的方法更好.几乎在一切时刻方法 P 都是最好,特别是,只要 $\lambda n=\lambda(n+2)$,我们总会有 $c_P(n)=c_Q(n)-1=c_R(n)-2$.但是,确实存在着无限多个 n,其时 Q 要优于 P 和 R(例如在 $n=2^{2^k}-2,k\geqslant 1$ 的场合),也确实存在着无限多个 n,其时 R 要比 P 与 Q 都好(例如当 $n=2^{2^{2^k}}-1$ 的场合).这些事实向我们暗示也许不存在胜过方法 R 的方法,按照那种意义,我们也许可以下结论:方法 R 是"最佳的".然而,这种想法却被下列事实所粉碎了.

事实 3 若 X 是任一标准表示方案,则必定存在着优于 X 的另一标准表示方案 Y,使对于只除掉一个值的所有 n 值,都满足关系式 $c_Y(n)\leqslant c_X(n)$.

[证] 本证法的大致思路是要选择一个长度为 c 的序列,然后用无限多个,相应长度为 $c+1,c+2,c+3\cdots$ 的序列去替代它,因 $2^{-c}=2^{-c-1}+2^{-c-2}+2^{-c-3}+\cdots$.

为了把证法写得更合乎常规,我们可以假定 X 是无冗余的.对 $k>1$,令 a_k 是能使 $c_X(n)=k$ 的那些 n 的个数,并令 j 为能使 $a_j>0$ 的最小的 j.设 $b_j=a_j-1,b_k=a_k+1$(对一切 $k>j$).则必存在一个唯一的非减函数 $c_Y(n)$,它恰有 b_k 个能满足 $c_Y(n)=k$ 的 n 值,此函数并能满足关系式 $2^{-c_Y(0)}+2^{-c_Y(1)}+\cdots=2^{-c_X(0)}+2^{-c_X(1)}+\cdots$.由事实 1 与事实 2,存在着一个具有成本函数 c_Y 的无冗余表示法.再进一步,容易

看出，除了独一无二的 n 值可使 $c_Y(n)=j$ 以及 $c_X(n+1)>j$ 之外，对其他所有的 n 值都有 $c_Y(n)\leqslant c_X(n)$； 此外，凡是 $j+1<c_X(n-1)<c_X(n)$ 的场合，都会有 $c_Y(n)<c_X(n)$. 证明完毕.

想要找出一个严格最佳方案是毫无希望的，因为只要我们反复应用在事实 3 的证明中所说的办法，将会得到一族越来越好的方案，其中的每一个都优于前一个. 然而，事实表明，没有一种方案能显著优于我们的递归方案 R，尽管事实 3 向我们保证可以进行逐步改良.

事实 4 设 Λn 为由下式

$$\Lambda n = \lambda n + \lambda\lambda n + \lambda\lambda\lambda n + \cdots$$

定义的函数，则每一个与一种表示方法相对应的成本函数 $c(n)$ 将能使无限多个 n 满足

$$c(n) > \Lambda n + \lambda\lambda^* n.$$

[证]令 $d(n)=\Lambda n+\lambda\lambda^* n$；我们将证明和式 $\sum_{n\geqslant 0}2^{-d(n)}$ 发散，有此即足以完成证明，因若对一切 $n\geqslant m$，$c(n)\leqslant d(n)$ 时，我们将有可能成立的不等式

$$1 = \sum_{n\geqslant 0}2^{-c(n)} \geqslant \sum_{0\leqslant n<m}(2^{-c(n)} - 2^{-d(n)}) + \sum_{n\geqslant 0}2^{-d(n)}.$$

一般地说，如果对任意函数 g，$f(n)=\Lambda n+g(\lambda^* n)$，则我们有

$$\sum_{n\geqslant 0}\frac{1}{2^{f(n)}} = \sum_{n\geqslant 0}\frac{1}{2^{g(n)}},$$

由于左边可以写作

$$\sum_{m\geqslant 0}\frac{1}{2^{g(m)}}\sum_{\lambda^* n=m}\frac{1}{2^{\Lambda n}}$$

并对 $m\geqslant 1$，有

$$\sum_{\lambda^* n=m}\frac{1}{2^{\Lambda n}} = \sum_{\lambda^*\lambda n=m-1}\frac{1}{2^{\lambda n+\Lambda\lambda n}} = \sum_{\lambda^* k=m-1}\frac{1}{2^{k+\Lambda k}}\sum_{\lambda n=k}1 = \sum_{\lambda^* k=m-1}\frac{1}{2^{\Lambda k}}.$$

因而当且仅当和式 $\sum_{n\geqslant 0}2^{-\Lambda n}$ 发散时，和式 $\sum_{n\geqslant 0}2^{-d(n)}$ 发散，而实际情况确是如此. 证明完毕.

利用同样的证明技巧，我们可以证明，对任何固定的表示方案，任何固定的 m，实际上将无限多次出现以下不等式.

$$c(n) > \Lambda n + \Lambda\lambda^* n + \Lambda\lambda^*\lambda^* n + \cdots \Lambda(\lambda^*)^m n + \lambda(\lambda^*)^{m+1} n.$$

怎样才能改进方案 R 以使我们获得极其接近于这一下界的“最终”方案？我们将进行说明，以此结束本节. 序列 $R(n)$ 由 $1^{\lambda^* n}0$ 开始，这是一个长度为 $\lambda^* n+1$ 的序列，其作用是识别 $\lambda^* n$，紧接着它的是一个长度为 Λn 的序列，其作用是一旦 $\lambda^* n$ 已知时，用以标志 n 的特性. 在此种意义下，R 方案是先从利用一元方案猜测 $\lambda^* n$ 入手的，而我们业已知道可以干得更好一些. 让我们利用 R 方案来决定 $\lambda^* n$，然后像以前一样来决定 n. 我们将此种办法称为 RR 方案. 新的成本函数将是

$$c_{RR}(n) = \Lambda n + c_R(\lambda^* n) = \Lambda n + \Lambda\lambda^* n + \lambda^*\lambda^* n + 1,$$

而 n 的 RR 表示将由序列 $1^{\lambda^*\lambda^* n}0$ 开始.

但是，请等一等！让我们从猜测 $\lambda^*\lambda^* n$ 开始，于是我们将得到一个 RRR 方案，其成本函数为

$$\begin{aligned} c_{RRR}(n) &= \Lambda n + \Lambda\lambda^* n + c_R(\lambda^*\lambda^* n) \\ &= \Lambda n + \Lambda\lambda^* n + \Lambda\lambda^*\lambda^* n + \lambda^*\lambda^*\lambda^* n + 1. \end{aligned}$$

R^{m+1} 方案的成本函数等于

$$c(n) = \Lambda n + \Lambda\lambda^* n + \cdots + \Lambda(\lambda^*)^m n + (\lambda^*)^{m+1} n + 1,$$

而这一上界几乎同我们的下界完全一样.

致　　谢

Philip Davis[4]写过一本趣味盎然的入门书，介绍了大数的一些初等性质. 表示方案的问题原先由 Levenshtein[11]探讨过，他提出了 R 方案，并且证明了存在着无限多个 n，可使 $c(n) > \Lambda n - \lambda^* n$；他也讨论了多于两个符号的表示方案. 表示问题也曾独立地由 Elias[5]，Even 和 Rodeh[6]进行过研究；Bentley 与 Yao[2]曾提出无界搜索的猜数游戏. 杰姆·鲍伊斯(Jim Boyce)与戴维·富克斯(David Fuchs)

在与本文作者最近的一次谈话中提出了 P 方案与 Q 方案.

在文献〔8〕中讨论了远较本文大得多的数.利用那篇文章中的记法,我们将有 $\lambda^*(2\uparrow\uparrow m)=m+1$,以及 $\lambda^*(2\uparrow\uparrow\uparrow m)=(2\uparrow\uparrow\uparrow(m-1))+1$,它提示我们:我们已推导出的上界与下界毕竟并不是靠得很近;在我们与超自然数打交道时,引进函数 $\lambda^{**}n$ 与 $\lambda^{***}n$ 等也许有助于弄清楚:究竟什么才算是"最佳"表示方案.

一位古代中国数学家徐岳(约公元200年左右在世)曾探讨过大数记法,其中的单位有"万"$(=10^4)$,"亿"$(=10^8)$,"兆"$(=10^{16})$,"京"$(=10^{32})$.请参阅他的一本有趣著作《数术纪遗》(见〔12〕的87页).我谨向董云梅(Tung Yun-Mei,译音)表示谢意,她为我提供了这篇参考文献.

在本文所讨论的表示问题与算法复杂性的信息论概念(最早由苏联的列文(L. A. Levin)与阿根廷的卡伊丁(G. J. Chaitin)独立地进行过探讨)之间有着一种美妙的联系.某种函数 $l(n)$(列文称之为 $KP(n)$,见文献〔7〕;而卡伊丁则称之为 $H(n)$,见文献〔3〕)具有以下两条性质:(a) 存在着一种自然数的表示方案,它满足条件1,并具有成本 $l(n)$.(b) 对每一种能满足条件1且具有成本 $c(n)$ 的表示方案,存在着一个常数 C,使满足关系式 $l(n)\leqslant c(n)+C$.直观地看,$l(n)$ 表示某一算法的长度,而此种算法是计算 n 的"最简单"描述.从计算数学观点看.此种函数 $l(n)$ 不能计算,但它可以从上面进行"半计算",这个意思是指,如果 $l(n)$ 当真比一个给定数 m 来得小,则我们可用有限时间证明这个事实.如果 $c(n)$ 是满足条件1与2的任意表示方案的成本方案,则存在一个能使 $\lambda n+l(\lambda n)\leqslant c(n)+C$ 成立的常数 C.另外,存在一个常数 C_0 以使得 $|\lambda n-l(\lambda n)-\max_{1\leqslant k\leqslant n}l(k)|\leqslant C_0$.于是,我们给定的作法将在 $\max_{1\leqslant k\leqslant n}l(k)$ 上提供界限.对这些文献材料,我应感谢 Péter Gács.

参考文献

1 Archimedes. 1956. The Sand Reckoner. In *The World of Mathematics*, vol. 1, ed.

James R. Newman, pp. 420-31. New York: Simon and Shuster.

2 Bentley, J. L. and Yao, A. C. 1976. An almost optimal algorithm for unbounded searching. *Inf. Proc. Letters* 5: 82—87.

3 Chaitin, Gregory J. 1975. A theory of program size formally identical to information theory. *Journal of the ACM* 22: 329—340.

4 Davis, Philip J. 1966. *The Lore of Large Numbers.* New Mathematical Library, vol. 6, New York: Random House.

5 Elias, P. 1975. Universal codeword sets and representations of the integers. *IEEE Trans. Inform. Theory* IT-21: 194—203.

6 Even, S. and Rodeh, M. 1978. Economical encoding of commas between strings. *Comm. of the ACM* 21: 315—317.

7 Gács, Péter. 1974. On the symmetry of algorithmic information. *Soviet Math. Doklady* 15, 5: 1477—1480, 15, 6: v.

8 Knuth, D. E. 1976. Mathematics and computer science: coping with finiteness. *Science* 194: 1235—1242.

9 Kraft, L. G. 1949. A device for quantizing, grouping and coding amplitude modulated pulses. M. S. thesis, Electrical Eng. Dept., Mass. Inst. of Technology.

10 Kronecker, L. 1893. Remark in a lecture at the Berlin scientific congress, 1886: "Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk." Quoted by H. Weber in *Math. Annalen* 43: 15.

11 Levenshtein, V. E. 1968. On the redundancy and delay of decodable coding of natural numbers. *Systems Theory Research* 20: 149—155.

12 Needham, Joseph. 1959. *Science and Civilisation in China* 3 Cambridge: Cambridge University Press.

答案与解法

1 阿基米德的最大数字应读作：One septendecyllion trevigintyllion quattuorvigintyllion quinvigintyllion sexvigintyllion septenvigintyllion octovigintyllion trigintyllion untrigintyllion quattuortrigintyllion septentrigintyllion octotrigintyllion novemtrigintyllion quadragintyllion duoquadragintyllion trequadragintyllion octoquadragintyllion novemquadragintyllion quinquagintyllion quattuorquinquagintyllion.（对这个数来说，他定的名称要短得多，但新命名法在许多其

他场合要比他的办法更为优越).

顺便提一句,大数的美国式叫法所用的拉丁文很蹩脚,譬如说,sexdecillion 应叫做"sedecillion",novemdecillion 应叫做"undevigintillion",所以人们搞不清楚美国式命名法究竟能够外推到多远.

2　一个 latinlatinlatinbyllionyllionyllionyllion 相当于

$$10^{2^{(10^{2^{(10^{2^{100000000000000002}}+2)}}+2)}}$$

为了简短起见,我们可以把它叫做一个 umptyllion.

3　R 方案的迭代译解过程如下:设输入序列为 $s_1s_2s_3\cdots$,其中每一个 s_i 都是 0 或 1. 设 k 为已读入的输入位数,l 为要输入之数的长度,m 为实现递归过程时的栈的深度,n 为返回的答数. 下面的递归方法具有 k 个初始零值,只利用唯一的全局变元,它是与定义直接对应的.

过程 R

置 $k \leftarrow k+1$.

如 $s_k=0$,置 $n \leftarrow 0$,退出过程 R.

否则,调用 R(递归地).

然后置 $l \leftarrow n$, $n \leftarrow 1$.

若 $l>0$,反复安置 $k \leftarrow k+1$, $n \leftarrow 2n+s_k$, $l \leftarrow l-1$.

退出过程 R.

通过机械转换,这一递归式程序可以转变为下列纯粹的迭代式程序:

置 $k \leftarrow 1$,以及 $m \leftarrow 0$.

在 $s_k=1$ 时,反复地安置 $m \leftarrow m+1$ 以及 $k \leftarrow k+1$.

置 $n \leftarrow 0$.

在 $m>0$ 时,反复地执行下列操作:

　　置 $m \leftarrow m-1$, $l \leftarrow n$, $n \leftarrow 1$.

　　在 $l>0$ 时,反复地安置 $k \leftarrow k+1$, $n \leftarrow 2n+s_k$, $l \leftarrow l-1$.

与此相应的编码过程可以类似地处理.

计数图论学家及其计数对象

● 滑铁卢大学

罗纳德·里德(Ronald Read)

什么是图？提起“图”这个词，大多数人就会想起图1中所示的那种商人用的曲线图，或者是一条光滑的曲线，如图2所示，它反映了某种数学函数的性质. 但是对于越来越多的数学家来说，这个词所

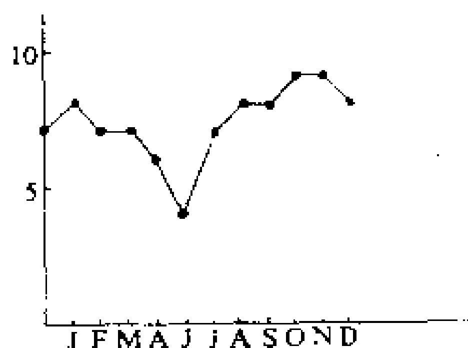


图1

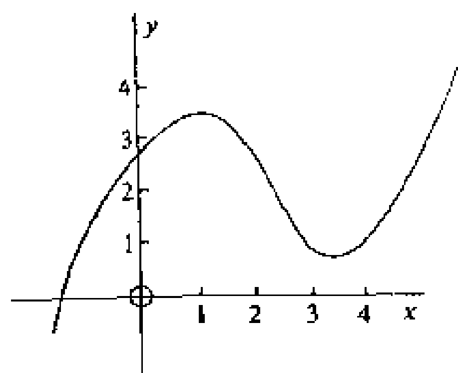


图2

代表的意思就截然不同了. 这就是人们所关心的，在某种意义上说是数学的一个分支，称为图论. 当人们（在这个意义上）说到图时，脑子里就形成了由一些点，或称结点，以及连结它们的直线或曲线所组成的图形，这些线常常被称作“边”. 如图3给出的就是这种例子.（本文所要讨论的正是这一类图.）这样的对象就是人们走在街道上会联想

到的网络(这个名词对于图论学家来说可能稍许还有些其他的特殊涵义).

图论　研究这些图及其性质的理论——起初是拓朴学的一个分支，事实上，它们曾被认为是“拓朴学的杂碎”，但它现在已变得受人重视(如果说以前不是的话)，被认为相当重要了，人们发现它的应用贯穿于数学与其他科学领域中的许多问题之中.

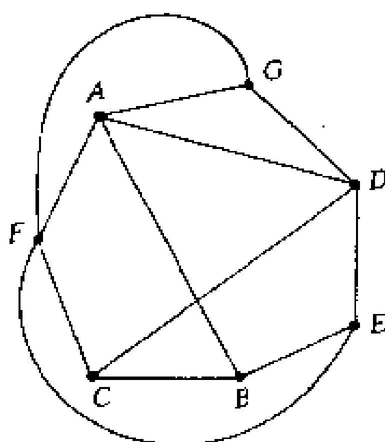


图　3

为什么一门初看起来并不比铅笔在纸上随意乱画更好一些的学科却会具有重大的作用?我们通过观察周围的事物来回答这个问题，看看日常生活中究竟有多少重要的事情本质上是图的问题或是能用图表示的问题.例如，出现在人们脑海里的是形形色色的通讯系统，电话网络(电话线把许多城市联络起来)，公路，铁路或航空线，等等.在这些例子中，对象本身就有着图的形状，例如线，路，等等.这些都是图形中的边，把一些结点(如上例中的城市)成对地连结起来.

我们同样可以把图论用到那些本身并不像图的事物之中，通过构造一个图来概括它的某些重要特性.例如，请你考虑一下集会中的一群人.我们可以这样构思：把每个人当作一个结点，当两个人相互熟悉时，就把他们用一条线连结起来.这样构成的图如图 4 所示.我们可以从这个图形中看出许多东西;例如，我们认为图 4a 的集会简

直沉闷之至；图 4b 的集会比较好，但是看来有“小集团”；图 4c 则是“野”得很.

(a)　　(b)　　(c)



图 4

值得注意的是，在平面上边与边之间相互不交叉的图有可能画不出，图 3 就是一例. 正因如此，我们才把每个结点画成小圆圈，从而不至于使那些类似于 AB、CD 的交叉点误为结点. 也有些图形没有这种交叉，这类图形称为“平面图”，我们将在下文中再次遇到.

作为图论中一个实际问题，我们假设有一个破坏者或是特务，他想破坏一个复杂的电话网络. 他想知道应该切断的最少电线数，或是应该炸掉最少的交换器，以便完全切断某些关键地点的一切通讯联络. 这个问题涉及到图的“连通性”. 我们还可以在每条边上标上数，来进一步深化图的概念，在这种场合下我们就得到了所谓的“网络”.



图 5

图 5 就是一例. 这些数可以是两点间的距离，也可以是两个地方之间的货物运费，或者是其他什么的. 可见，网络可以用来描述广泛领域中的内容，而运筹学中一些重要而不断发展的领域也联系着网络及其特性.

正是由于这个原因，几乎图论的所有研究都有其实际的应用，至少是潜在有用. 然而也有例外. 一些图论学家躲在图论大厦的黑暗角落，仔细地避开现实世界的光线，他们忙忙碌碌在解决的问题，离开实际很遥远. 这些人热衷于图的计数，所关心的问题是："如此这般的图一共有多少种？"这些图论计数的爱好者们花费许多时间在清点图形的个数，或者寻找更好的点数方法. 这是一种奇特的工作，但并非毫无乐趣.

"计数图论学家"到底是什么样的人，他们究竟在数些什么东西？我希望在本文结束时至少能给读者一个初步的回答.

凯莱(Cayley)和树的计数

图论计数的先驱是亚瑟 · 凯莱(Arthur Cayley)勋爵，他从 1857 年开始提出和解决了许多关于"树"的计数问题. 图论学家所定义的树是一种连通(整个图成为一个整块)而没有回路的图. 为了理解什么是一个回路，请考虑一张公路网络图. 这里所指的一个回路，就是

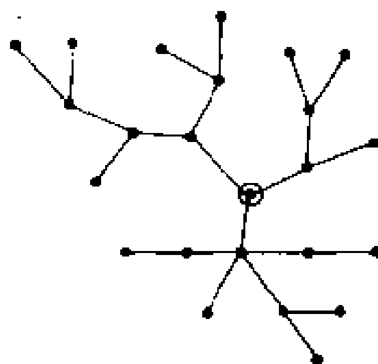


图 6

从公路的起始点开始走，在经过每个结点不多于一次的情况下回到起始点的路径. 如果把回到起始点这个条件去掉，那就是图论学家所谓的"路径". 另一种关于树的定义是：任意给定两结点间只有唯一路径的一个图. 图 3 所示的图不是树，因为它有回路 *ABCDEFGA*，图 6 所示的图是树，读者很容易明白为什么这些图可以称为树.

树中边的数目总是比结点数少 1（**问题 1**：请证实这一点）. 因此关于树计数的最基本的问题是"包含 n 个结点的树有多少?"这就是一个凯莱所解决的问题.

图论计数问题中，弄清究竟数些什么东西是十分重要的. 当我们问"这样的和诸如此类的不同图形有多少"时，我们尤其需要知道我们所指的两个图象的相同和不同是什么意思，这随着问题的不同而异. 通常，如果我们能把每个结点标上整数 1，2，3，…，P，而两张图相对应的结点连接方式相同，则我们认为这两个图是相同的（或用专业名词"同构"）.

因此，图 7 中所示的两个图，初看是完全不同的，实际上它们却是同构的. 图中显示了给两个图标上标号的一种方法以说明这一点.（**问题 2**：图 8 显示了含 10 个结点的四张图，其中三张是同构的，另

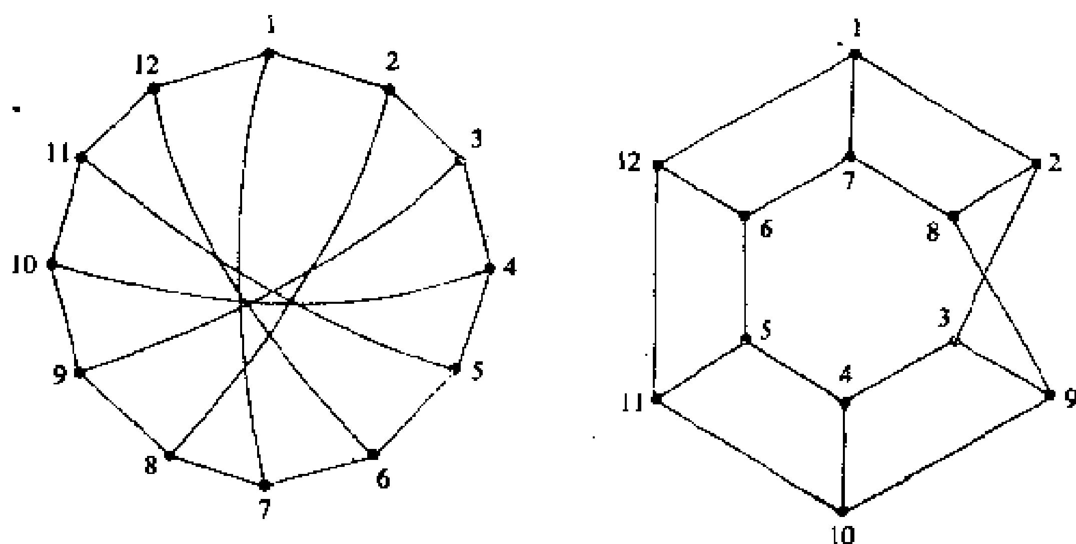


图 7

外一张与它们不同，谁能指出它？）

因此，当我们问“有多少不同的图”时，我们必须说明我们所指的不同是什么意思．通常它就意味着“不同构”，但这并不是必要的；我们还有其他的想法．

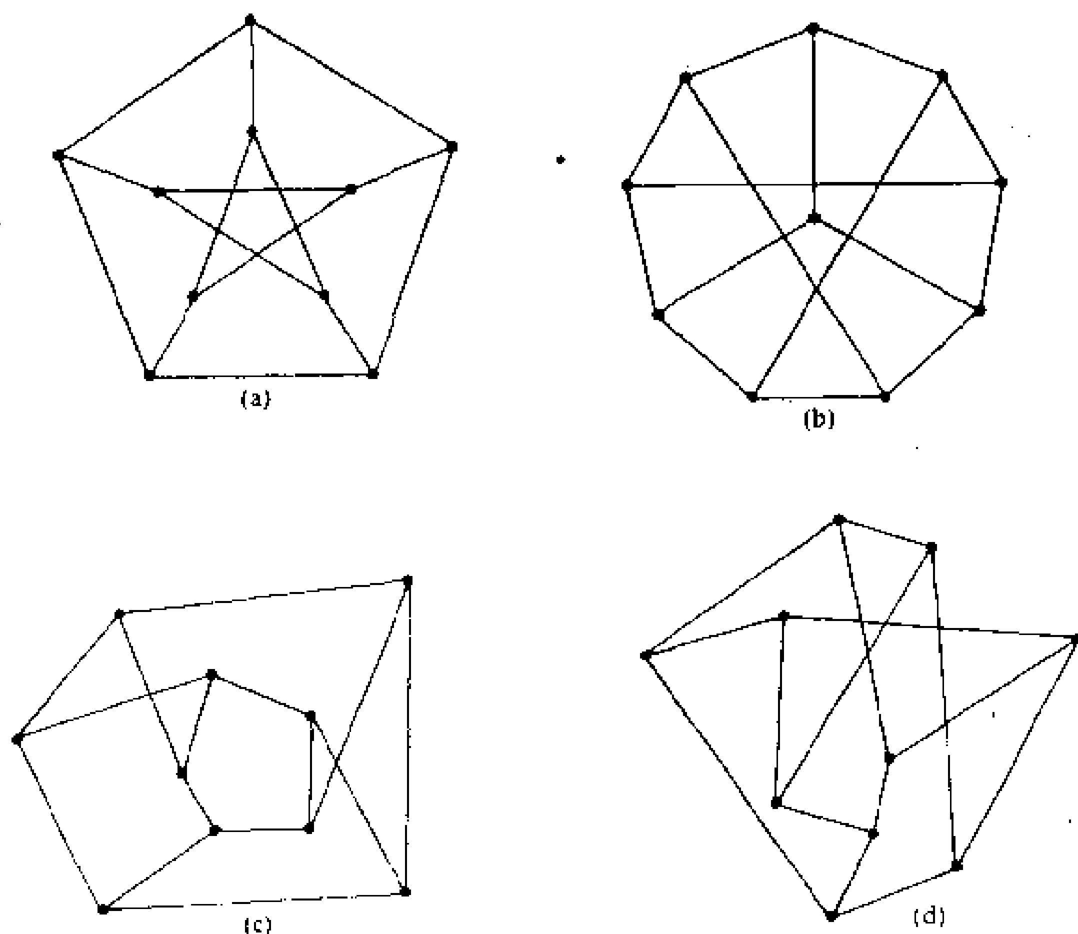


图 8

二元树的计数

现在，让我们来仔细看看一个比较简单的问题，以体会一下图论计数中的几种方法．我们将计数二元树．一个二元树首先是一个有根

的树.这意思就是树的一个结点，我们称之为根，可以用某种方式与其他结点区别开来.在画一个有根的树时，根可用圆圈圈开来（如图6所示），或者把根画在图的底部.然后，在二元树的每一结点上，至多引两条向上伸的边，它们总是从根往上长的.如果正好是两条，那么一条往左伸，一条往右伸.如果只有一条，则或者向左伸，或者向右伸，这两种可能性被认为是不同的.如果没有枝，那这个树在这个结点处就不再向外发展，这个结点就是树的端点或叶.图9显示了含有4个结点的所有二元树.请注意，由于向左伸和向右伸的区别是十分重要的——是定义二元树的一部分——所以这些树是完全不同的.如果我们不去强调这种区别，那么前面两个树就将是同构的，图中其他几对树一样也将变成同构了.

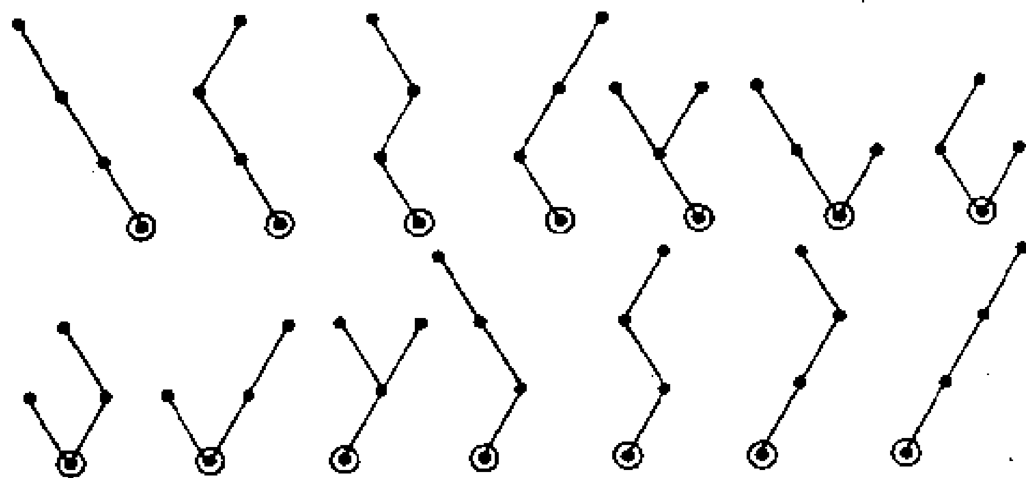


图 9

在对二元树进行计数时，我们注意到对于给定的任意二元树，例如含 $n+1$ 个结点的二元树，从它的根部起总是有两条向上的分枝——一条向左，另一条向右，它们被定义为两个子树——一个左子树和一个右子树，如图10所示.如果用 B_n 表示含 n 个结点的二元树的数目，让我们假设已经知道 $B_1, B_2, \cdots, B_n$ 的数值，为了确定 B_{n+1} 这个数，我们考虑如何构成含 $n+1$ 个结点的二元树.

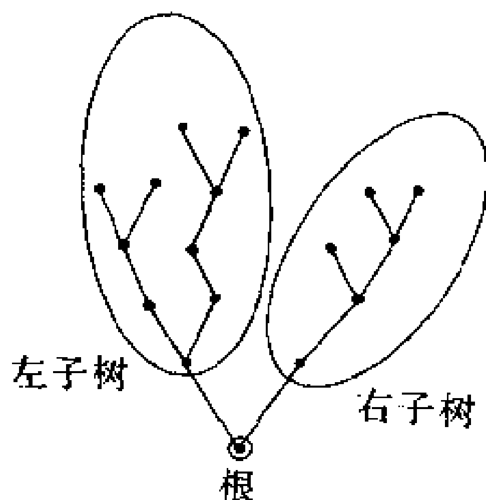


图 10

这就相当于挑选两个子树作为左、右子树(它们总共含有 n 个结点),把它们的根连接成一个新的结点——这样整个树的根就构成了.那么,我们有多少种选择方法呢?如果左、右树分别含 l 个结点和 r 个结点,那我们将有 $B_l \cdot B_r$ 种选择方法;因为含 l 个结点的 B_l 个树中的任何一个结点都能与含 r 个结点的 B_r 个树中的任意一个相组合.

如果我们将 l 与 r 赋予不同的值(注意 $l+r=n$),并把结果相加,那我们将得到含 $n+1$ 个结点的二元树个数.但请注意!我们已经假设从根出发有两条分枝,并且任何时候都正好有两个子树;但有些情况也可能只有一个子树.值得庆幸的是我们用个小技巧就能避免这类困难.如果我们要数的一个树只有一个子树,那我们就假设它有两个子树,但其中另一个子树是"空树"——它并无任何结点!

运用这个方法,我们就可以认为每一个二元树都给出两个子树.因为"空树"只有一个,故我们定义 $B_0=1$,并允许 0 为 l 或 r 可取的值.在这个意义上,我们得到:

1. $B_{n+1}=B_0B_n+B_1B_{n-1}+B_2B_{n-2}+\cdots+B_{n-1}B_1+B_nB_0$,

这样,当 $B_0,B_1,\cdots,B_n$ 都已知时,我们就可以求得 B_{n+1} 的值.

读者们也许会关心下列数字的验证：

n	0	1	2	3	4	5	6	7	8	9	10
B	1	1	2	5	14	42	132	429	1430	4862	16796

这些数 B_n，是很有名气的卡塔朗数，它们在大量组合问题中经常出现．这些数除了是含 n 个结点的二元树之个数外，B_n 还是把 $n+2$ 边形剖分成三角形的方法数，也是将圆周上 $2n$ 个点用不交叉的弦连成对的方法数，又是将 $n+1$ 个字符正确组成括号的方法数．关于这些数的一些有趣的论述及其性质的文章，马丁·加德纳曾发表于《科学美国人》1976 年 6 月号上．

用母函数的形式来给出问题的解答是计数问题经常用到的一个有效办法．为了从卡塔朗数 $\{B_n\}$ 之类序列中推导出一个"母函数"，我们把它们作为一个无穷幂级数的系数，故卡塔朗数的母函数是：

$$B(x) = B_0 + B_1x + B_2x^2 + B_3x^3 + B_4x^4 + \cdots$$

现在不难看出，1 式的右边已成为 $B(x)$ 与其自身相乘的展开式中 x^n 的系数，也就是 $x[B(x)]^2$ 中 x^{n-1} 项的系数．而 1 式的左边是 $B(x)$ 中 x^{n+1} 项的系数，因此推导出下列结果：

2. $B(x)=1+x[B(x)]^2$.

右边加上 1，表示的是"空树"，这在 1 式中是没有包含的．

等式 2 是 $B(x)$ 的二次方程．可解出

$$B(x) = \frac{1-\sqrt{1-4x}}{2x},$$

并从此经过冗长但并不困难的代数运算，我们得到关于卡塔朗数的一个明确表达式：

$$B_n = \frac{(2n)!}{(n+1)!n!}.$$

前面的方程显示了有根树计数的一个常用技巧．即先把树根移开，观察所得到的子树，这时会出现许多情况，因为有许多种不同类

型的树等待我们点数.我们也许会得到多于两个的子树,——甚至是可变量,这些树可能是有序的(如同刚刚作出的左右之区别),也可能是无序的.但不管如何,每个子树都是有根的(这个结点与原始根相连),只要保证它们与原来的树具有相同的类型,我们都有指望能推导出某种递归方式来获得所需要的数字.几乎所有树的计数都是遵循这种模式的.

凯莱曾经计数过一些树,其中主要是我们所说的"普通"有根树,这种树对一个结点汇集多少边是没有限制的,至于这些边的次序及图在平面上的描绘也是无关紧要的.他指出,如果 T_n 是含 n 个结点的这类树的个数,那就有下列相当奇怪的等式:

$$T_1x + T_2x^2 + T_3x^3 + \cdots$$
$$= x(1-x)^{-T_1}(1-x^2)^{-T_2}(1-x^3)^{-T_3}\cdots.$$

同前面的一样,如果对从 1 到 n 这 n 个数,T_n 的值都已知,那么把这些数代入等式的右边,我们就可以从等式的左边算出序列的下一个数.

凯莱也计数过一些"无根树".这个问题就困难多了,譬方说,因为没有根就不能把图分割开来,也就不能把问题分成各个小问题来解决,如我们在前面所做过的那样.凯莱还计数过一些不同类型的化合物,从而为图的计数理论的实际应用开辟了一个新天地,赋予其特殊的用途.这些化合物的结构式(如烷烃——石蜡及其衍生物)实质上就是一些特殊类型的树图.

一般的图的计数

对一般的图来说,其边数与结点数之间并不存在如同树一样的关系;一个含有 p 个结点的图可以有从 0 到 $p(p-1)/2$ 之间的各种不同边数,(0 是"空图"的边数,$p(p-1)/2$ 是每两点都相连的完全图的边数).于是问题就得这样来问:"总共有多少结点数为 p,边数为 q 的不同的图?"其中 p 与 q 是给定数.

这个问题凯莱不曾想要解决，即使有所尝试，也未必能成功，因为当时解决它的工具还没有产生.这个工具便是由 G. 波利亚(G. Pólya)于 1938 年发表的一条定理(见〔9〕)，而在他的几篇更早的文章里实已有所蕴含.波利亚定理是解决许多组合问题的基础.虽然在这里我们不能仔细讨论，但了解一下其大概内容也是有用的，它在解决图论计数问题中大有用处.

波利亚定理

相当数量的组合问题可以归结为“放物入盒”问题.让我们假设有一组不同的盒子，在每个盒子里我们放入一件东西——通常是个图案，进而每幅图案都以一个非负整数为其值.对每个盒子来说，都存在一个图案集，其中有一幅图案要放进盒子(通常的情况是每个盒子可供选择的图案是一样的，但这并非必需).在每只盒子里我们选择恰好一个图案放入，最后就形成了一个构形——盒子及其所含的图案.这个“构形”的值被定义为放入该盒子内的各图案的值之和.

我们首先要提问的是“如已给定如何放图入盒的全部有关要求，那么在已知整个构形的值时，有多少种组成此构形的方法?”作为本问题的一个例子，假设我们要把 21 根蜡烛插入一个矩形生日蛋糕上，蛋糕的表面被分成了六块，如图 11 所示。我们约定可以在每一

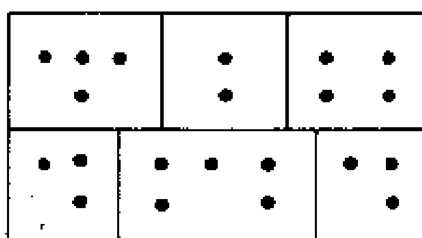


图 11

部分上插入任意根蜡烛，而这些蜡烛在每块蛋糕中的精确位置都是

不重要的，重要的是各个蛋糕块中插入蜡烛的根数．在这个例子中一个图案包含一定数的蜡烛(可以根数为 0)．而图案的值就是蜡烛数．问题是如何寻找整个值为 21 的构形数．

这个问题并非特别困难，所以我们马上着手把它搞得难些．我们假设所有的盒子并不是完全不同，故对盒子作某些特定的重新排列并不能导致与其他排列相区别．

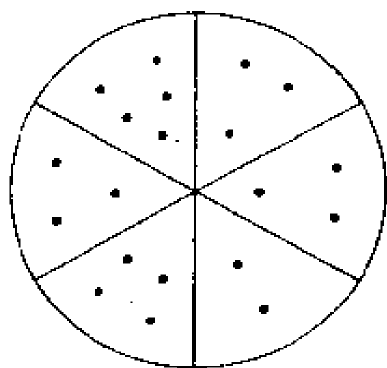


图 12

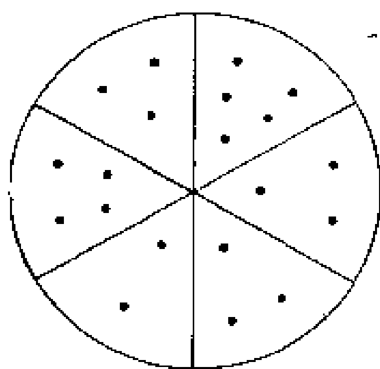


图 13

例如，如果用一只圆形蛋糕代替矩形蛋糕，同样分成六等分，那把图 12、图 13 看作是不同构形的作法就是不现实的，因为把前者顺时针旋转 60 度就能完全覆盖在后者上．从解决问题的目的着眼，仅仅是方向不同的两种构形，我们认为是同构的．对一般的对象来说，我们假定存在一个特定的置换集合(构成一个群)，并且认为：任意两种构形如果可以通过其中某一置换由一个转变为另一个，则它们是等价的．当我们问及不同构形的数目时，我们所指的是不等价的构形数．对于圆形生日蛋糕，这个群是由旋转 60 度的倍数所形成的不同置换组成的．请注意，图 14 所示的图与图 13 的图是不同的，因为无论旋转多少度，它们都不可能达到完全一致．

本文中，我们不可能仔细论述波利亚定理；实际上我们甚至不打算正式叙述它．我们仅仅能意识到它可以为解决上面所勾划的一般

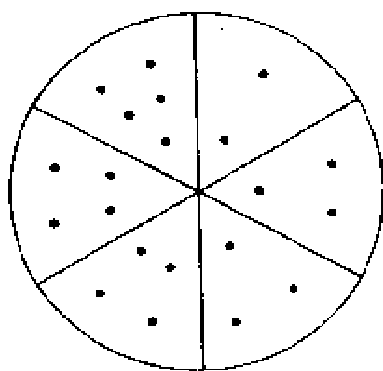


图 14

性问题提供一条求解途径. 如果给出问题的数据(给出放什么图案入盒的条件以及它们的值),并给出刻划两种构形是否等价的置换群,我们就可以应用波利亚定理来找出答案,说得更确切些是一组答案,因为波利亚定理提供的是一个给定任意值求构形数的母函数.

我们如何把图的计数问题归结为"放物入盒"问题? 我们按下面的步骤做. 考虑有 p 个结点的图,分别标上 $1,2,3,\cdots,p$,那就有 $p(p-1)/2$ 对结点,而每对结点都有可能组成图的一条边. 把这些结点对当作问题中的"盒子",每只盒子只允许放两种图案,"无边"图案的值为"0","有边"图案的值为 1. 在每只盒子里放入相应的图案后,显然就形成了一张图——如果盒子(i,j)内是"有边"图案,那节点 i,j 是相连的——而构形(图)的数值就是它包含的边数. 以这种方式形成的图是有标号的图,因为每个结点都在集合$\{1,2,\cdots,p\}$中选取不同的标号. 这并不是我们所要的. 如果两个图同构,那我们不会因为其结点标号不同而认为它们是不同的. 我们必须承认这个事实,由于不是所有的结点标号都不同,所以,结点对——盒子也是如此. 对结点的标号作置换自然导致盒子间的置换. 如果我们把标有 1、2 的结点分别改作 5、9,那(1,2)盒子就变成(5,9)盒子,等等. 结点标号的每一次置换,都引起盒子的置换,而盒子的一系列置换就形成了问题

中相应的群. 运用波利亚定理可以得到问题的答案.

正是运用这种方法，现在图论领域中最知名的人物之一，密执安大学的弗兰克·哈拉里(Frank Harary)于1955年开始图论计数. 用来解决问题的母函数相当复杂——甚至有点可怕——但能用来计算图的数目，尤其是在计算机的帮助下，C·A·金(C. A. King)和E·M·帕尔麦(E. M. Palmer)已经计算出 p 从1到24的所有图的数目. 24个结点的所有非同构图的数目是下面比天文数字还大的数：

$$195704906302078447922174862416726$$
$$256004122075267063365754368.$$

(这算是真正有用的信息吗？你是怎么看的？)

这一理论也成功地应用于各种变相的图的计数，尤其是"有方向的图"或(通常简称)有向图，它与一般图的不同之处就在于它的每条边都有方向(用箭头表示)，或向这个方向，或向另一方向，也可能同时向两个方向. 图15中就给出了一个典型的有向图. 有向图的计数仅仅在细节问题上与普通图不同，用波利亚定理也能给出答案. 这个问题由哈拉里在计数一般的图的同一篇论文中所解决([5]).

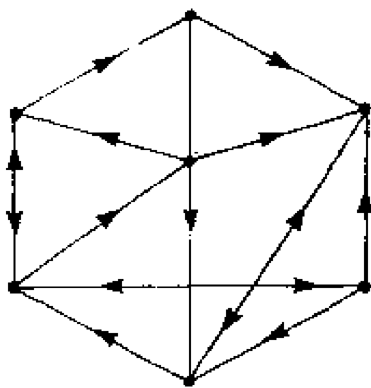


图 15

尽管波利亚定理是很有力的，但它也不能解决图论计数中的一切问题. 能够证明这一点的是自补图的计数. 图 G 的"补图"定义为

与 G 具有相同结点，G 有的边它没有，而 G 中没有的边它都有的图. 图 16 在同一组结点中设置了一个图(用粗边表示)及其补图(用细边表示)，向人们显示了图及其补图合在一起恰好包含了含 p 个结点的图可能有的全部 $p(p-1)/2$ 条边. 有可能出现补图与原图同构的情况，即所谓自补图. 最简单的自补图包含 4 个结点，3 条边，如图 17 所示. 对含 5 个结点的图来说，共有两种不同的自补图，如图 18 所示. 这种简单例子的数目总是可以通过反复试探获得，但是当结点数上升为 8 时，反复试探的方法就再也不能寻找出究竟有几个自补图了. (**问题** 3：当结点数为 6 或 7 时，有多少个自补图?)我们所需要的是一个理论结果，它能告诉我们结点为任意值 p 时这类图的数目.

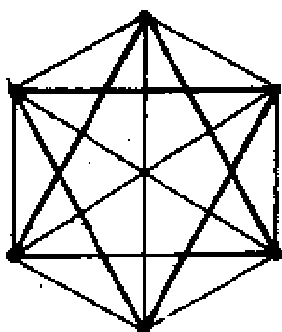


图 16

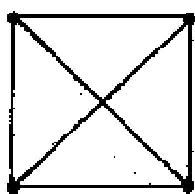


图 17

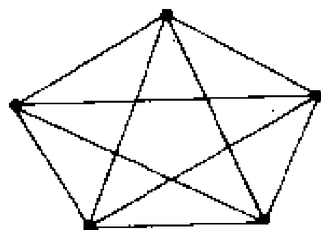


图 18

这类问题，需要一种比波利亚定理更有效的理论，因为它的复杂性更强. 图与补图的区别在于我们把"边"与"非边"这两个字眼进行了交换，这样便从一个名词过渡到另一个名词. 因此，我们得到了一种不同于一般的问题——为了判断两种构形是否等价，我们允许盒子的置换，同样也允许图案的置换. 由此，我们获得两种置换群——一种是盒子的置换群，一种是图案的置换群——这样，所有事物就变得更为复杂了. 对付这类普遍而具典型性问题的方法最早由丹麦数学家德·布鲁因(N. G. deBruijn)在 1959 年发表的论文中给出[1]，此后不久就解决了自补图的计数(参看[10]). 有向自补图的定义与

其对应图的概念很类似. 图 19 显示了一个有向图与它的自补图同构,这类有向自补图的计数是由一般自补图发展而成的.

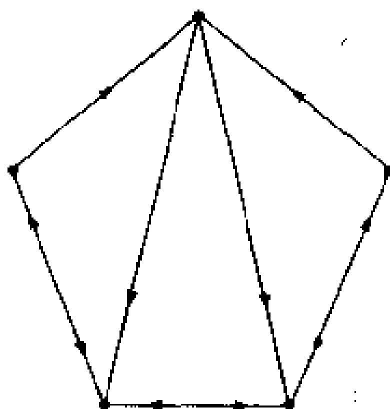


图 19

从上述结论中,冒出了一个非常奇怪的事实:结点为偶数的有向自补图数目恰好等于结点为其两倍的自补图的数目. 如此直接的关系,在直觉上使人强烈地意识到有向图与普通图之间必定存在着某种简单的联系;一定存在较容易的方法断定上述数目是相同的,并能说明为什么. 或许确实如此,但真是这样的话,至今却尚未发现.

随着时间的推移,许多更为有效的计数定理产生了——有的是对波利亚与德 · 布鲁因定理的推广,有的则开辟出完全不同的研究方向. 其中,哈拉里(Harary)与帕尔麦(Palmer)的"幂群计数定理"尤为重要,但它实在太复杂,不能在此述说. 读者如需进一步了解,可去读这两位作者的著作《图的计数》[7]——这一图论专门分支中一本包罗万象的书.

莱德菲德(J. H. Redfield)的奇异情况

现在我们应当对出现在图的计数舞台上一个不平常的人物简单说上几句. 也许更确切地应该说是"重新出现"……. 早在 60 年代前

期，弗兰克·哈拉里就提请他的同伴们注意一篇引起他自己注意的、由他的一个学生提供的文章[11]．文章的标题是："成群递减的分布理论"，是由一位至今未被人所知的数学家 J.H.莱德菲德于 1927 年所发表的．在这篇文章中（正如哈拉里所发现的，也是我们在惊讶中证实的）莱德菲德给出了一条与波利亚定理等效的定理；他已经涉足于图的计数（至少是部分——不过很清楚他也许走得更远）；他超前使用了几项"最新"的图论中的发现．值得高兴的是，他还证明了一条有效的新定理．自从发现我几年前在博士论文中提出的一个主要论点即是莱德菲德已得出的一条结论时开始，我就记住了莱德菲德定理是被重新发现的这件事（有关在同一组结点上几张图迭合而成的图的计数问题）．这样看来，三十几年来，图论计数者及其他一些学者苦心钻研的理论、解决问题的方法及其证明早已被莱德菲德在 1927 年所证实！然而他的著作却完全没有引起注意．怎么会有这样的事情？

这并不是因为他的文章登在一本毫无影响的杂志上，事实上它是登在众所周知并享有很高声誉的《美国数学杂志》上．也许几个平时足以引人注意的因素恰好都不具备，莱德菲德在他自己的研究领域中不为人所知——事实上这篇文章也确实是他唯一的数学著作；而文章的标题又没有向人们暗示文章内容；他在文章中所用的记号又是一些奇异的符号，虽然印在纸上，却没有提醒人们文章究竟写些什么．或许最有说服力的因素是："时机尚未成熟"．

莱德菲德主要并不是一个数学家——他的研究领域是（或者曾经是）语言学．尽管如此，他唯一的一篇数学论文却是十分有价值的．假如他的文章早些被重视，那图论计数理论的发展轨迹将毫无疑问地与现在很不一样．当然，由于重新被发现，莱德菲德的著作已经激发了一些新进展．尤其是澳大利亚纽卡瑟（Newcastle）大学的罗宾孙（R. W. Robinson），已经采用并改进了莱德菲德几个观点，而且解决了几个以前认为很难对付的问题．

平面图的计数

我们已经知道平面图的含义，就是在同一平面上边与边互不相交的图. 现在让我们看看有关它们的几个计数问题. 自然要问："具有 p 个结点，q 条边的平面图究竟有多少个？"但这个问题太难了. 让我们考虑"三阶的平面图"，其定义为：每个结点正好汇集 3 条边的图. 在三阶平面图中，结点数 p 与边数 q 的关系是：$2q=3p$（**问题 4**：为什么？），因此只要一个数就能确定图的大小——即结点数（必须为偶数）. 我们希望这将使问题简单化，但是看上去还是很难. 这是平面图的一种典型状况，其计数问题往往出奇的困难，但当一个问题能够被解决时，它的解决方法却通常比一般图简单.

如果我们进一步给问题加上约束条件，那就很能说明这一点，譬方说，提出问题："具有固定的哈密尔顿回路的三阶平面图有多少？"所谓图的哈密尔顿回路是这样一个回路——可把它称作"周游路径"——它访问了图中的每一结点. "固定"的含义是哈密尔顿回路一旦画出之后就在平面上固定下来了，只有剩余的边才可以通过不同的位置改变形成我们所要计数的不同对象. 现在让我们看看图 20 中显示的图，其中不在哈密尔顿回路上的边用细线表示. 这些细线要末都在回路里面，要末都在回路外面. 因为，如果它们与回路相交，则

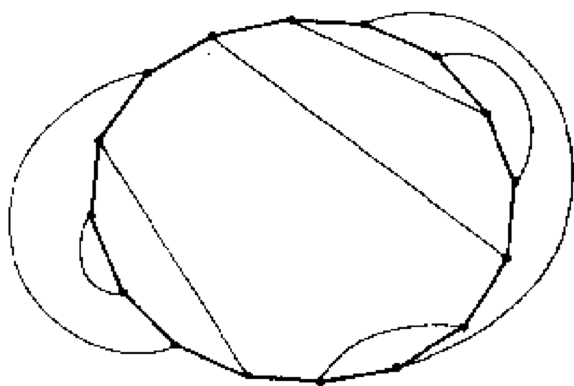


图 20

就不是平面图了.进一步看,每一个结点都恰是一条细边的端点.由此我们可以用简洁的话把这些图的计数问题归结为:"给定一个有 $2n$ 个顶点的多边形,那么有多少方法能用不交叉的弦把成对端点连接在多边形之内,并把其余点用不交叉的"弦"(因为缺少一个更好的字眼)连在多边形外?"

这个问题被滑铁卢大学的著名教授 W·T·塔特(W. T. Tutte)用一种极为优雅的方法解决了,他在平面图领域作过深入的研究,并在平面图计数中做过大量工作.让我们简短地看看他的结论和证明过程(详见[12]).

首先,我们考虑如果把所有的结点都连在多边形之内,那就得到了一个很简单的问题.实际上,它与许多问题一样,其答案就是卡塔朗数的序列,上面也已提到过.因此,如果我们约定已知的 $2r$ 个结点是内部弦端点,那么成对地连接这些点的方法数是卡塔朗数:

$$\frac{(2r)!}{r!(r+1)!}.$$

对于剩下的 $2s$ 个组成外部弦的结点(满足 $r+s=n$)也有类似的结果.

由于每一种内部弦的安排方法都可以配置一种外部弦的安排方法,所以该把这两个数字相乘.我们还应该乘以选择这两组结点的方法数,这也就是在全部 $2n$ 个结点中选择 $2r$ 个结点作为第一组的方法数,即组合数

$$C_{2n}^{2r}=\frac{(2n)!}{(2r)!(2s)!}.$$

取遍各种不同的 r、s 值,对这些乘积求和,就得到:

$$\sum_{r+s=n}\frac{(2n)!}{(2r)!(2s)!}\cdot\frac{(2r)!}{r!(r+1)!}\cdot\frac{(2s)!}{s!(s+1)!};$$

也就是:

$$\sum_{r+s=n}\frac{(2n)!}{r!(r+1)!s!(s+1)!}.$$

通过一些简化过程,最后得到十分简单的表达公式:

$$\frac{(2n)!(2n+2)!}{n!(n+1)!^2(n+2)!},$$

它便是我们所需求的图的数目.

当 $n=2$ 时,上式的值为 10. 也许读者会感到奇怪,当结点为 4 时,居然存在那么多种可能性,但看一下图 21 就可知道为何如此. 这里的回路是正方形,由于不允许旋转,图(1)、图(2)是不同的. 进一步看,虽然图(3)和图(5)是同构的(即使考虑边的颜色),但因为画在平面上的方式不同,故也认为它们是不同的.

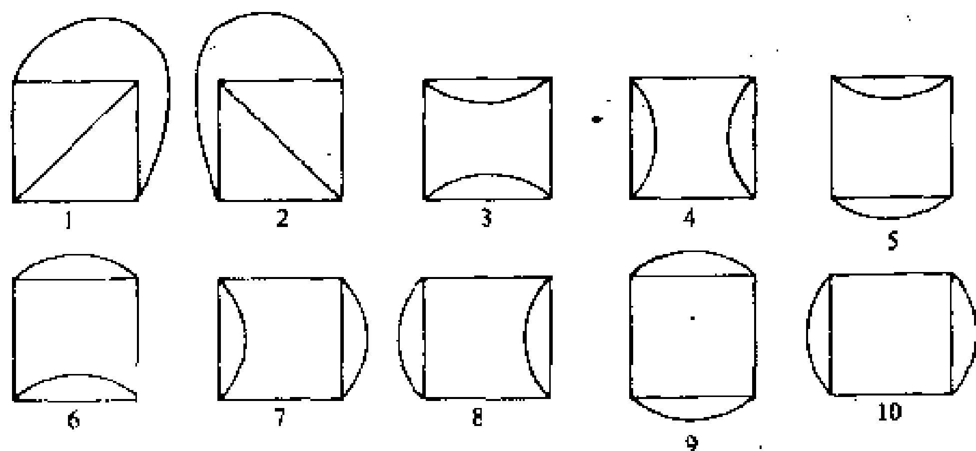


图 21

当然,在两结点间存在两条边的情况也是允许的. 这类"双边"情况,有时允许,有时不允许,要根据问题的性质而定. 在这里,我们约定不发生"双边"情况似乎更合理. 但如果我们允许这样做了,我们会遇到计数问题中无定性的一个例子. 这个小小的变化带给我们的却是一个困难得多的问题——就我所知,迄今尚未解决.

在图论计数理论中有许多未解决的问题,其中有几个问题在可能解决的边缘徘徊,有的似乎只需要在目前的正确理论与技巧上再进一步就足以征服它们. 也有一些确实是棘手的,已知的一切手段都攻不动. 无论怎样,这都为读者提供了一个运用数学智谋的诱人的机

会，如果你喜欢干这类事情，它将足够你消磨许多时间.

问题的答案

问题 1 树的边数总是比结点数少 1.

我们可以由单个结点开始画树，并一条一条地增加边，可以看到每增加一条新的边总是把一个老结点与一个新结点连接起来. 在每一步中，我们在已有的边数上增 1，在已有的结点数上增 1，因此，从开始画树时（边数为 0，结点数为 1）起，这个结论就是正确的，画完树时，也是正确的.

问题 2 图 8 中的图.

图(c)是不同的. 可以从图 22 给出的标号看出其余的图是同构的. 图(c)包含一个含 4 个结点(*ABCD*)的回路，而其余图的回路所含

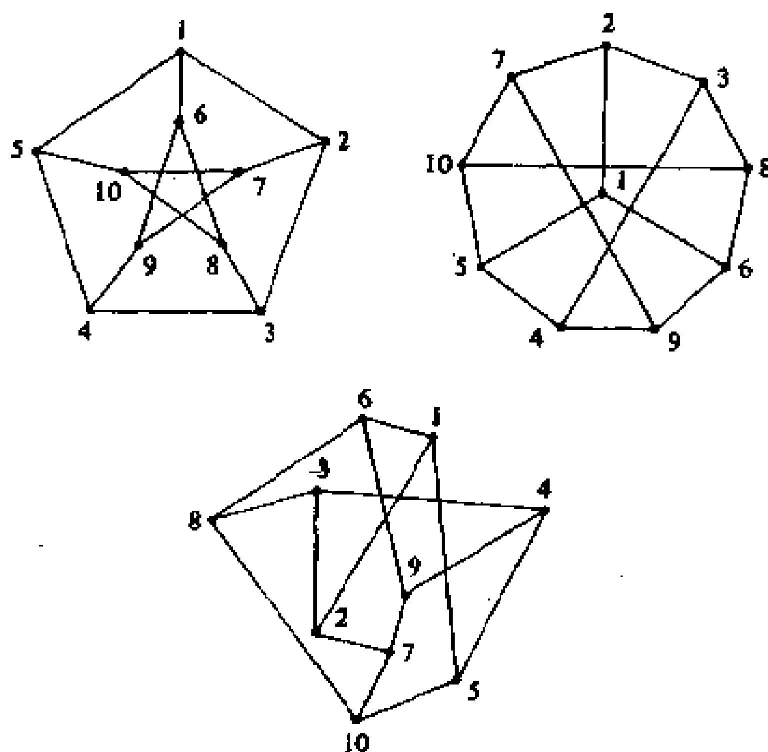


图 22

结点都不少于 5. 故(c)不可能与其他几个同构.

问题 3 包含 6 或 7 个结点的自补图.

根本没有. 一个图与它的补图总共含有 $p(p-1)/2$ 条边,进一步,因为它们是同构的,它们必定含有相同的边数. 故 $p(p-1)/2$ 定是偶数,这对 $p=6$ 或 7 是不成立的. 自身互补图仅仅当 $p(p-1)/4$ 为整数时才存在,也就是当 p 除以 4 余 0 或 1 时.

问题 4 在三阶图中 $2q=3p$.

因为每条边有两个"端",于是整个图共有 $2q$ 个"端",而每个结点汇集了 3 个"端",故 $2q=3p$.

参考文献

除了正文中已提到的论文外,以下文献中还包括了普遍感兴趣的材料.

1 deBruijn, N. G. Generalization of Pólya's fundamental theorem in enumerative combinatorial analysis. *Indag. Math.* 21:59—69.

2 Cayley, A. 1857. On the theory of the analytical forms called trees. *Phil. Mag.* 13: 172—176.

3 Cayley, A. 1874. On the mathematical theory of isomers. *Phil. Mag.* 47: 444—446.

4 Cayley, A. 1875. On the analytical forms called trees, with applications to the theory of chemical compounds. *Rep. Brit. Ass.* 257—305.

5 Harary, F. 1955. The number of linear, directed, rooted and connected graphs. *Trans. Amer. Math. Soc.* 78: 445—463.

6 Harary, F. 1969. *Graph Theory*. Reading, Massachusetts: Addison-Wesley.

7 Harary, F. and Palmer, E. M. 1973. Graphical enumeration. New York: Academic Press.

8 King, C. A. and Palmer, E. M. Calculation of the number of graphs of order $p=1$ (1)24. (unpublished.)

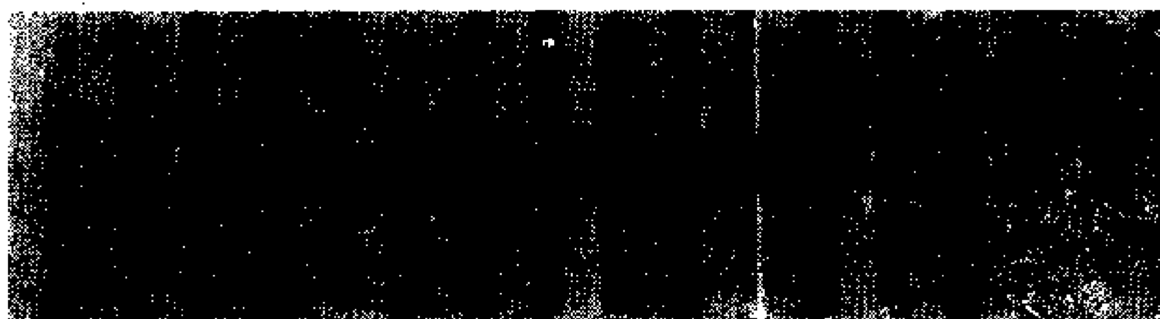
9 Pólya, G. 1937. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.* 68: 145—254.

10 Read, R. C. 1963. On the number of self-complementary graphs and digraphs. *J.*

London Math. Soc. 38: 99—104.

11 Redfield, J. H. 1927. The theory of group-reduced distributions. *Amer. J. Math.* 49: 433—455.

12 Tutte, W. T. 1976. Hamiltonian circuits. Colloquio Internazionale sulle Teorie Combinatorie. *Atti de Convegni Lincei.* 17: 194—199.



● 贝尔实验室

□ N·J·A·斯隆纳(N. J. A. Sloane)

这篇文章打算写成一篇导论，以报道晚近十年来密码学所取得的一系列激动人心的进展. 戴维·卡恩(David Kahn)的有趣著作《破译者》[1]出版于1967年〔29〕，不幸的是它刚出版，IBM的罗锡弗(Lucifer)编码方案〔11〕、〔20〕、〔51〕就出台了. 后者激发了我将对之进行描述的一系列进展.

1967年以前，涉及密码学的数学论文相当枯燥(有人说是故意要写得乏味，旨在不鼓励这方面的研究). 但在读完此文之后，我想你们肯定会同意，新进展当真是振奋人心的. 对我说来，它们是一些长期以来通信论进展中的一些迷人概念(由于它们中间绝大多数都不是我的发明，因而我不必为了对它们如此热心支持而感到惭愧)，为了避免在正文中不时出现一些最高级的赞美之辞，让我干脆在这里一次性地把本文所根据的六篇经典的精采论文介绍一下：

Shannon(1949,〔49〕)，

Feistel,Notz 与 Smith(1975,〔13〕)，

Wyner(1975,〔56〕)，

---

[1] 译者注：《破译者》是美国密码协会主席戴维·卡恩的名著，被誉为密码学的"圣经". 此书已由群众出版社于1982年译成中文，但被删去不少内容.

Diffie 与 Hellman(1976,〔5〕),

Rivest,Shamir 与 Adelman(1978,〔44〕),

Merkle 与 Hellman(1978,〔38〕).

也许除了我对事物有自己的看法之外,本文没有什么新东西. 不过,我将在文中提及如此众多的编码办法,谅来绝大多数读者会发现其中有着自己所不熟悉的内容.

顺便提一句,我必须着重强调指出,我是一位不保密的密码学家,我对这方面的保密内容一无所知. 幸运的是,那些懂得保密的读者已经有誓言的约束不能加以谈论,因而即使我讲得不对,他们也不能纠正我的错误. 我将要加以叙述的一切编码办法都是在公开文献中发表过的,然而文中所表达的一些看法则属于我自己,既不是参考文献中所提到的那些论文作者,也不是贝尔系统的专家们.

本文分为五部分,共同点都是沿导线传输信息的问题(见图 1).



图 1

让我们假定,我们可以沿着导线发送点与划,或者,说得更数学化一

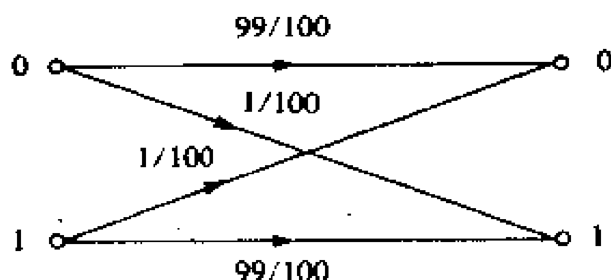


图 2

一个二元对称信道. 0 被发送时,一百次中有 99 次能收到 0,但一百次中有一次收到的是 1,发送 1 时情况也与此类似. 所以这一信道的错误概率是 1/100.

点，发送 0 与 1. 当我们发送 0 时，通常在远处终端出现的也是 0，但有时(由于噪声原因)，收到的是 1. 反之，发送 1 时，收到的通常为 1，但有时也会收到 0. 这种导线有时称为二元对称信道，可用图 2 来描述.

要发送的数据业已转换成一串 0 与 1(也许通过计算机来产生)，我们的问题如下：

尽可能用导线传输更多的信息，要求更快、更可靠地传递，并切实防止窃听.

我将描述这个问题的五种不同解决办法，其条件一个比一个更困难(见图 3). 第一种情形(见第 1 节)是，信道只是图 1 所示的那种简单情况，解决办法是利用纠错编码法. 第二部分是，我们有着同样的有噪声信道，但出现了第二条可疑的导线(也可参看下面的图 10)：有了一个窃听者！开始时，我们不妨假定窃听者使用的设备极为低劣，我们将说明这是很容易对付的(第 2 节). 本文的第三部分将要谈到窃听者具有第一流装备时应怎样对付. 这就是传统密码术的用武之地. 人们将利用一种发送者与接收者都了解的密钥，但坏家伙对它却是一无所知(第 3 节). 第四部分讲到一种特殊情况，此时我们当真把信息传给坏家伙要他为我们传输. 但是我们仍有办法挫败他(第 4 节). 最后一部分描述了所谓的公开密钥体制：对第 3 节所提到的一些办法作了重大改进以使得发送者不需要了解密钥(第 5 节). 在所有的情况下我们都能挫败坏家伙，安全地发送我们的信息，使他无从得知其内容(或无法进行篡改——这在第 4 节里会讲到).

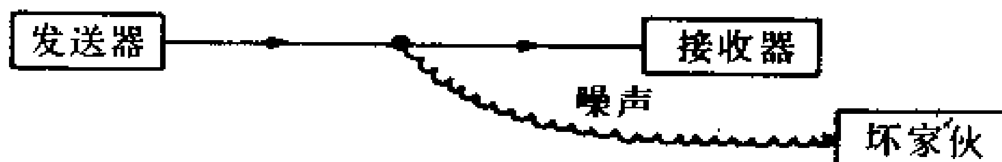
第 1 节
纠　错　码

问题是怎样才能经过图 1 与图 2 的信道尽可能迅速与可靠地传输信息. 编码家的解决办法是不让 0 与 1 的任何旧序列发送出去，而

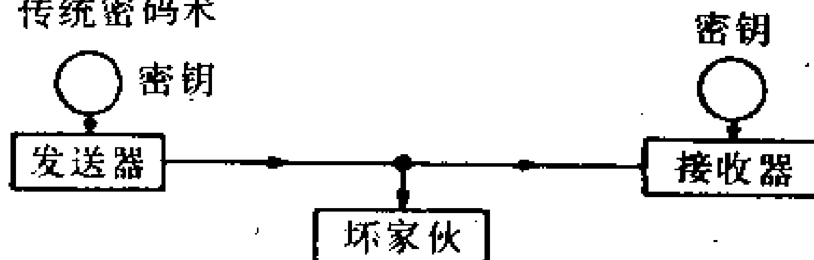
I 纠错码



II 有人窃听的信道



III 传统密码术



IV 察觉欺诈行为的编码



V 公开密钥体制
无密钥!

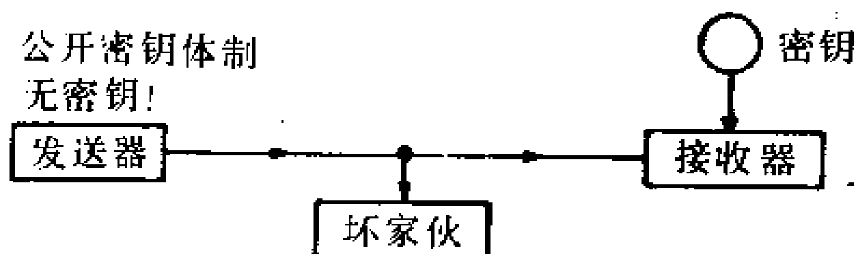


图 3

本文中所讨论的五种不同通讯系统.

只准发送人们称之为“码字”的特定序列. 所有准予发送的码字清单称为码典;发送者与接收者都了解它,他们正进行通力合作,力图消灭传输差错.

信息将通过码字来代表,当一个特定的信息要进行传输时,发送出去的是其对应的码字. 接收端出现一个码字(或许是它经过畸变之后的形式)时,译码者必须决定发送人发出的究竟是哪一个码字并从而理解其所代表的信息. 图 4 表明一个非常简单的编码实例: 只有两个码字,即00000与11111,长度都是 5,分别对应于信息“是”与“非”. 设信息为“是”,传输出去的码字为00000,收到的(由于存在干扰噪声)却是01000. 译码者懂得,在信道中差错并不很多(见图 2),于是他断定,传送出去的码字必定是00000,信息为“是”. 一般地说,译码者所应选取的码字,是在某种意义上与收到的序列最为“靠近”的.

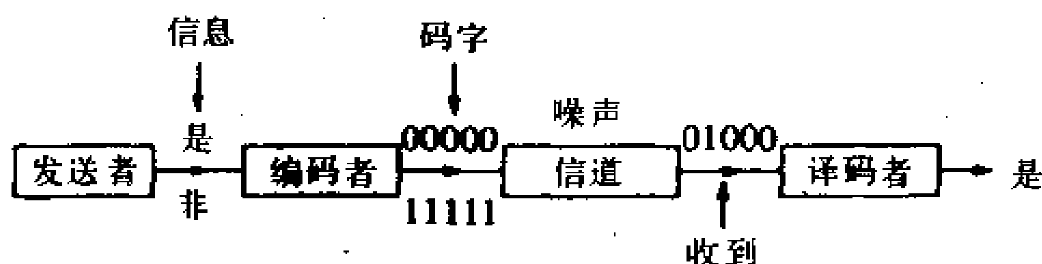


图 4

一种简单纠错码的运用,共有两个码字:00000 与 11111. 收到的序列是01000,它被正确地译解为“是”.

为了说得更确切,让我们来定义序列之间的距离. 长度为 n 的两个序列 $u=(u_1,u_2,\cdots,u_n)$ 与 $v=(v_1,v_2,\cdots,v_n)$ 之间的汉明(Hamming)距离(简单地记为 $\text{dist}(u,v)$)就是它们不同的位数. 例如

$$\text{dist}(01000,00000)=1,$$

$$\text{dist}(01000,11111)=4.$$

现在我们就能授予译码者更准确的训令:他应该把收到的序列

译解为汉明距离最接近的那个码字. 后者是最有可能发送的码字. 图4说明了这个过程，我们可以看到在这个例子中译码者有能力改正一个错误. 而事实上这种编码方案甚至能改正任一对差错. 譬如说，发送的是00000，而收到的却是01100(第二位与第三位出了毛病)，然而01100仍然更接近于正确的码字而不是更接近于11111. 从另一方面说，这种编码方案显然纠正不了三个错误(如果收到的是11100，我们将把它错误地译解为11111)，因此，它是一种双差错纠正码.

理由是明白无误的：码字本身之间的汉明距离为5. 同样的论证对任何一种编码方案全都适用，最重要的参数是码字之间的最小汉明距离：

$$d = \min_{u \neq v} \operatorname{dist}(u, v),$$

对一切不同的码字 u, v 都来取. 这么一来，如果出现的差错不超过 $e = [(d-1)/2]$ ($[x]$ 表示不超过 x 的最大整数值)个，则收到的向量将比任何其他向量更接近于正确的码字，从而译码者即能作出正确的决定.

总而言之，在码字之间具有最小距离 d 的码将是一种能纠正 $e = [(d-1)/2]$ 个差错的纠错码. 因此，我们希望能够找到一种 d 很大(可纠正许多错误)而长度 n 却很小(以便节省传输时间)的编码办法.

以下给出的编码实例将比图4中谈到的更加有趣得多.

1. 偶数权重码 E_n，包含了所有字长为 n，且其中1的个数为偶数的二进位码字，也就是说，它的权重是偶数. 图5给出的是 E_4 码. 至于一般情况下的 E_n 码含有 2^{n-1} 个码字，而码字之间的最小距离为 $d=2$. 这种编码对纠正差错并无用处，虽然它能察觉一个错误. 我们将在第2节中再次讨论它.
2. 长度为7的汉明码. 它是具有几何构造的许多种编码方案之一. 阶数为 p 的射影平面(第4节中将重新出现这

0	0	0	0
0	0	1	1
0	1	0	1
1	0	0	1
0	1	1	0
1	0	1	0
1	1	0	0
1	1	1	1

图 5

E_4 码，包括了所有字长为 4，且其中 1 的个数为偶数的八个向量.

种几何对象）是具有 p^2+p+1 个点和 p^2+p+1 条直线的集合，它们的配置法是：每条直线上都有 $p+1$ 个点，每个点都有 $p+1$ 条直线通过. 图 6 给出的是阶数为 2 的射影平面.

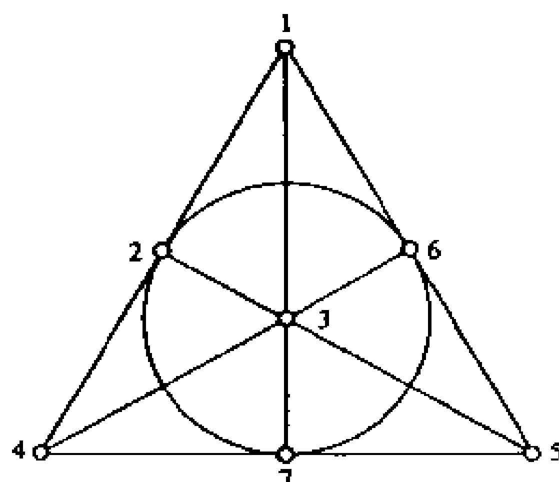


图 6

阶数为 2 的射影平面. 有七个点，分别标记为 1，2，…，7，与七条直线（其中的一条是曲线——即图上的圆）. 每条直线包含三点，每个点位于三条直线之上.

为了描述射影平面，并不是非要画图才行．作为作图方法的取代，人们只要安排一个元素为0与1的矩阵，说明每条直线上有哪几个点就行（这件事可以通过打电话交代清楚）．譬如说，图6所含的直线，其各点配置便是：

1,2,4
2,3,5
3,4,6
4,5,7
1,5,6
2,6,7
1,3,7

我们也可用变相办法，通过一个矩阵来描述图形．这时，矩阵的行表示直线，列表示点（见图7）．

点

线	1	2	3	4	5	6	7
	1	1	0	1	0	0	0
	0	1	1	0	1	0	0
	0	0	1	1	0	1	0
	0	0	0	1	1	0	1
	1	0	0	0	1	1	0
	0	1	0	0	0	1	1
	1	0	1	0	0	0	1

图 7

元素为0与1的矩阵可用来描述图6的射影平面．例如，矩阵的第一行表明，有一条直线包含着点1,2,4.

我们所需要的汉明码包含着这一矩阵各行的求补(由0与1的互换而得出)以及零的码字.它在图8中给出,最小距离 $d=4$,并且是纠正单个差错的纠错码.

0	0	0	0	0	0	0
0	0	1	0	1	1	1
1	0	0	1	0	1	1
1	1	0	0	1	0	1
1	1	1	0	0	1	0
0	1	1	1	0	0	1
1	0	1	1	1	0	0
0	1	0	1	1	1	0

图 8

字长为7的汉明码,它有八个码字,可通过图7中各行的求补运算而得出.

我们已讲过的所有编码法都具有一种特殊性质,这使它们便于编码与译码.此种特性是:它们都是线性码.也就是说,如果把两个码字,逐位进行模2加法,则其结果也将是一个码字.例如,在图5中,第二码字与第三码字的模2和,即0011与0101的和是0110,仍然是码典中的一个码字.

一种线性码可通过它的所谓奇偶校验矩阵来简单明瞭地给出其定义.后者是具有下列性质的矩阵 H:当且仅当

$$Hu^T \equiv 0 (\mathrm{mod}\ 2)$$

时(此处的 T 表示转置运算),向量 $u=(u_1,\cdots,u_n)$ 是一个码字.为什么要称为奇偶校验矩阵,如想解释清楚,势必要花费很多笔墨,请读者自己去参阅文献[35]的第一章,这里就一概从略了.

图 5 那种编码法的奇偶校验矩阵是

$$[1111],$$

这是由于当且仅当

$$u_1 + u_2 + u_3 + u_4 \equiv 0 (\bmod 2)$$

时，(u_1, u_2, u_3, u_4) 是一个码字.

对图 8 的编码法来说，我们可以用

$$\begin{bmatrix} 1101000 \\ 0110100 \\ 0011010 \\ 0001101 \end{bmatrix}$$

作为一个奇偶校验矩阵.

在纠错码问世三十多年以来，人们已经发现了大量的好例子，目前，这一理论已经发展得极其壮大. 事实上，F. J. MacWilliams 与我刚刚写完了一部多达 760 页的巨著〔35〕. 其中有很多编码是通过巧妙选定的奇偶校验矩阵来定义的. 特别有威力的一族是所谓哥帕（Goppa）码. 可惜，有关它们的阐述需要用上一些有限域的知识，因此，下面的一段文章读者可以自由选读，不必勉强.

3. 字长 $n=2^m$ 的哥帕码. 在伽罗瓦（Galois）域 $GF(2^m)$ 上选取一个次数为 t 的不可约多项式 $G(z)$. 作下列奇偶校验矩阵

$$H = \begin{bmatrix} \frac{1}{G(\alpha_1)} \cdots\cdots \frac{1}{G(\alpha_n)} \\ \frac{\alpha_1}{G(\alpha_1)} \cdots\cdots \frac{\alpha_n}{G(\alpha_n)} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\alpha_1^{t-1}}{G(\alpha_1)} \cdots\cdots \frac{\alpha_n^{t-1}}{G(\alpha_n)} \end{bmatrix}, \quad (1)$$

此处 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是域 $GF(2^m)$ 的元素. 码字是能使 $Hu^T=0$ 的所有二进制向量 u. 这种编码的性质在图 9 中作了概括.

哥帕码（类似于 BCH 码、Reed—Muller 码以及另外一些我们没有

提到的编码）有一个很有效的译解算法——从接收到的序列找出最接近码字的一种代数过程. 不过，对绝大多数编码来说，情况决非如此，这是第 5 节 5e 中将要进行描述的公开密钥体制的一种基本构思. 有关纠错码的进一步信息，读者可以去参阅文献〔35〕.

哥帕码的主要性质：

字长 $n=2^m$

码字个数至少为 2^{n-mt}

最小距离 $d\geqslant 2t+1$

它是一种能纠正 t 个差错的编码

图 9

第 2 节
有人窃听的信道

我们将要讨论的信道要从图 1 变为图 10.

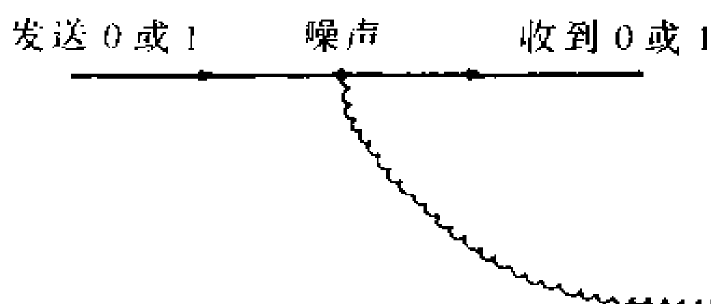


图 10

有人搭线窃听！

第二根导线是旁路，它一直通向坏家伙，此人正在窃听，什么话都逃不过他的耳朵. 于是问题的性质也就有所变动，我们的目标是尽快且尽量可靠地传输信息，并使窃听者所截获的内容越少越好.

在本节中我们将讨论的是利用低劣设备的蹩脚窃听装置，也就

是说，我们假定在搭出去的旁路上本身存在着噪声——窃听者处在图 2 中所示的二元对称信道的情况. 本节的主要参考文献是 Wyner 的著作〔56〕，也可参阅〔1〕，〔31〕，〔32〕，〔55〕. 最简单的情况是主线上不存在噪声——见图 11.

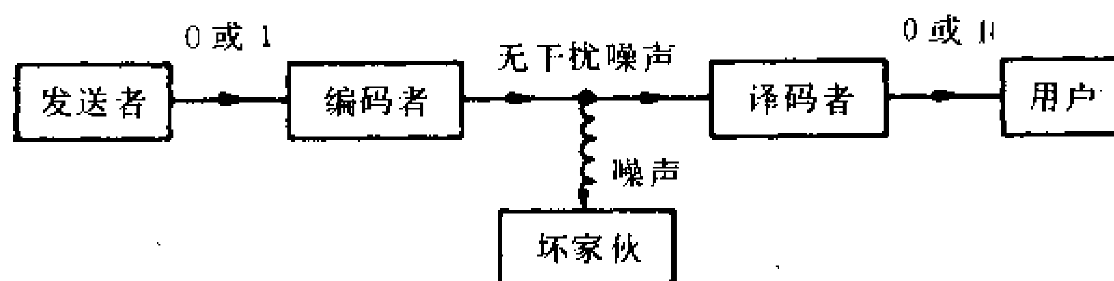


图 11

有窃听的信道的一种最简单情况，主线上不存在噪声，但旁路则是一种二元对称信道（见图 2），差错概率为 p_0.

这个问题有着一个很漂亮而令人惊讶的简明解法. 关键思路如下：

把 0 改编为一个长长的、随机选定的 0 与 1 的数字串，其中 1 的个数是偶数. 把 1 改编为一个长长的、随机选定的 0 与 1 的数字串，其中 1 的个数是奇数.

数字串的长度有意要搞得很长，以使得通向窃听者的旁线上足以引起几处差错. 譬如说，我们可以把 0 改编为

$$Y = 101110100000011111010100$$

（字长 24 位，其中 1 的个数为偶数的序列），但当它到达窃听人那里时，可能已改变为

$$Z = 100110100110011111010100.$$

即使窃听者了解编码办法（不过他当然无法知道选定的序列 Y 究竟

是啥东西),他还是被彻底打垮了.他明知 Z 是某一个序列 Y 的畸变形式,但他却根本不知道 Y 中究竟含有偶数个 1 还是奇数个 1.

反之,合法的收信人则能收到与发送时毫无差别的 Y,因此他只要点一点 Y 中 1 的个数就行了.他只要把 Y 中所有的各位(二进制位数)按模 2 求和即可.[1]

我想你会同意,这是一种异常聪明的想法.当然,它有一个明显的缺点——传输极其缓慢.信息的传输速率几乎是 0,这是因为要把一位数字传给收信人就得用上全部 Y 之故.

不过,这个缺点很容易修补.让我用另外一种方式描述.设 F^n 为字长 n 的一切二进制数向量的集合.我们把 F^n 划分成两个子集:子集 E_n(见第 1 节)包括了具有偶数个 1 的一切向量,余下来的向量(把它们称做 D_n)当然包括了奇数个 1.于是

$$F^n = E_n \cup D_n. \tag{2}$$

群论的语言在这里有用处.F^n 是一个群(在分量相加的意义下)而 E_n 是一个子群(请回忆,在第 1 节中我们曾注意到它是一个线性码).而且,D_n 是 E_n 的一个陪集或平移:

$$D_n = 100\cdots0 + E_n,$$

方程 2 说明了一个原理,即总是有可能把一个群划分为子群的陪集.我们的编码方案如下,把 F^n 划分为 $E_n \cup D_n$.在发送一个 0 时,改用 E_n 中的一个随机向量来代替;要发送一个 1 时,则从 D_n 中随便选个向量来代替.

现在我们可以来讲对付旁路窃听的普遍有效办法.选取一个良好的线性纠错码 C_1,它包含了 2^{n-k} 个字长为 n 的码字(见第 1 节).再把 F^n 划分为 2^k 个 C_1 的陪集,即

$$F = C_1 \cup C_2 \cup C_3 \cup \cdots \cup C_{2^k}.$$

把有可能发送的信息编号,从 1 号到 2^k 号.然后采取如下办法:

---

[1] 译者注:实际上即是计算 Y 中 1 的个数是奇数还是偶数,作者故意使用这种不通俗的说法,真令人有故弄玄虚之感.

把第 i 个信息改编为从陪集 C_i 中任意选取的一个随机向量.

对合法的收信人来说，很容易察知发送的究竟是什么信息，因为要决定一个向量究竟属于 C_1 的哪一个陪集是极其简单的(人们只要计算一下向量的分量就行，请参阅文献[35]的第 16 页)，可是窃听者却被击败了. 这种办法的传输速率是 k/n(为了给 2^k 个信息中的一个进行编码，需要发送 n 位)，进而，Wyner[56]能证明下面的结论.

定理 1 若 p_0 是窃听线上出现差错的概率，则人们有可能以低于

$$-p_0\log_2 p_0-(1-p_0)\log_2(1-p_0)$$

的任何速率传输信息而使窃听者对通信内容一无所知.

这个定理的证明需要引入一个称为"熵"或"不确定性的测度"的函数 H(见文献[16]的第 1 章，[36]的第 2 章). 它可以通过下面的方式来定义，设 X 是原始信息，Y 是编过码的信息，Z 是窃听者听到的 Y 的畸变信息，然后用 $H(X)$ 标志窃听者在听到 Z 之前所具有的关于 X 的不确定性，而 $H(X|Z)$ 标志他在听到 Z 之后所具有的关于 X 的不确定性. 显然

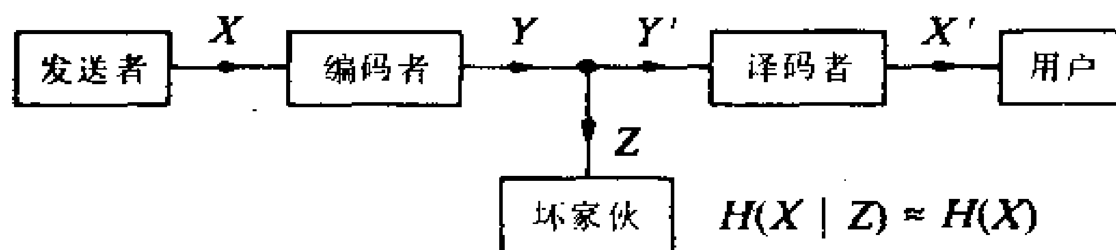


图 12

如果坏家伙在得知 Z 之后关于原始信息 X 的不确定性 $H(X|Z)$ 同他在架设窃听线路之前有关 X 的不确定性 $H(X)$ 实质上几乎一样，那么完全保密就有了保障.

$$H(X|Z) \leqslant H(X).$$

为了证明上述定理，人们只要能证明事实上成立

$$H(X|Z) \approx H(X)$$

就行了. 换句话说，窃听者徒劳无益地搭设了旁路，实际上了解不到什么情况(图 12). 他还不如躺在家里享福为好.

如要了解证法本身，读者可以参阅文献[56]. Wyner 也还分析了更一般的情况，即从发送到接收的主线上也存在着噪声的情况.

第 3 节
传 统 密 码 学

上一节我们轻易地击败了只有劣等装备的窃听者. 可是我们又将如何对付一位专家呢？他的设备竟是如此完善，以致于他能毫无畸变地听到所说的一切. 这就是需要应用传统密码学的情况. 我们先来讲利用密钥的编码办法，这时只有发送者与接收者了解密钥，而坏家伙则根本不了解(见图 13). 接下来，我们将在第 5 节中叙述一种新发现的编码方案，其特点是：只是接收者才需要知道密钥.

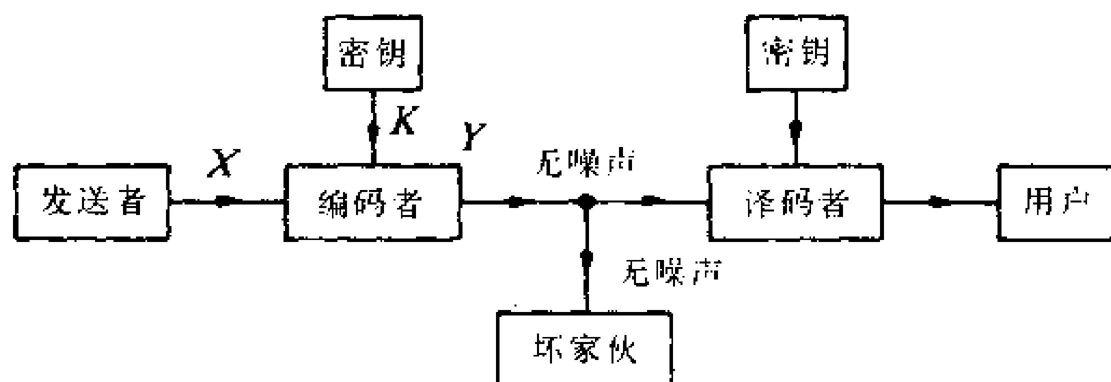


图 13

利用密钥的传统编码法. 只有发送者与接收者才知道密钥，而坏家伙则不知道.

3A 一次性密本

一切编码方案中，最简单与最安全的是一次密本(图 14).(对此项发明作出贡献的有几位学者，特别应提到 G. S. Vernam，他当时正在美国电报电话公司研究部工作，请参看文献[29]的第 13 章以及文献[54].)

作出一个长长的、0 与 1 的随机数序列，也许是通过多次抛掷一枚钱币而产生的，这便是密钥，然后把它穿孔在纸带上，并复制一份同时送给发信人与接收人. 发送者只须把这串数字逐位地与信息序列按模 2 求和，即可得出编码后的密文. 例如，如果信息是

$$\cdots\cdots 0100001101$$

而密钥序列是

$$\cdots\cdots 1101110011$$

则编码后的信息便是它们的和

$$\cdots\cdots 1001111110$$

(逐位按模 2 求和，没有进位). 在接收端，把收到的编码信息与密钥序列再一次按模 2 求和，于是原始信息就重新出现：

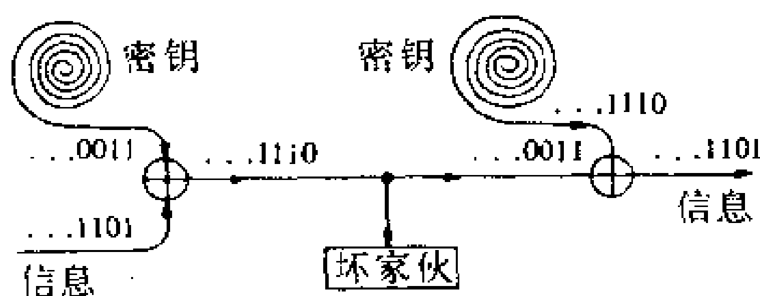


图 14

一次密本法. 对密钥与信息进行模 2 求和. 密钥序列与信息序列等长而且只使用一次.

编码后的信息……1001111110

加 密 钥 序 列……1101110011

得出原始信息……0100001101.

密钥序列仅仅使用一次并旋即毁去.这是一种完全不可破译的密码.因若令 X 表示信息序列,K 表示密钥数字串,

$$Y = X + K \tag{3}$$

表示编码后的信息.窃听的坏家伙虽然知道 Y,但由于一切不同的密钥数字串都是同样可能的,于是任何信息也有同样可能出现.因此他偷偷架设接听线路,实质上依然是毫无所获(见图 14).

这种办法的唯一缺点是:密钥与被传送的数据具有同等长度.尽管有这个缺点,在特别重要的信息传输中它仍然被广泛使用着.

然而对例行性的事情,人们还是想要一种密钥的容量较小的编码办法.设计一种优良编码方案的艺术是企图找到一种扩展密钥的方式,也就是说,取一个小容量的密钥以之作为种子,产生一个长得多的密钥数字串.目前已经知道有许多种方法来做到这一点(例如可以参看文献[8],[15],[29]),但此处我们只想提一下利用移位寄存器的某些技术.

3B 线性反馈移位寄存器

最早、最简单与最脆弱的办法是利用一个线性反馈移位寄存器的输出作为密钥数字串(图 15).

若移位寄存器有 m 级,则有可能通过适当选取反馈的连结方式,以产生周期等于 2^m-1 的输出序列(例如,可参看文献[22]或[34]).值得指出的是,这些输出序列是一个名为 Reed—Muller 纠错码(见文献[35]的第 406 页)的码字.如 m 很大(譬如说 100),这一周期就会长得像个天文数字($2^{100}-1\approx10^{30}$).

即便如此,这种密码仍然极易破译.为了测验密码的威力,必须假定坏家伙已了解到编码的算法(在这种情况下是移位寄存器)并已得出相当数量的破译对子:

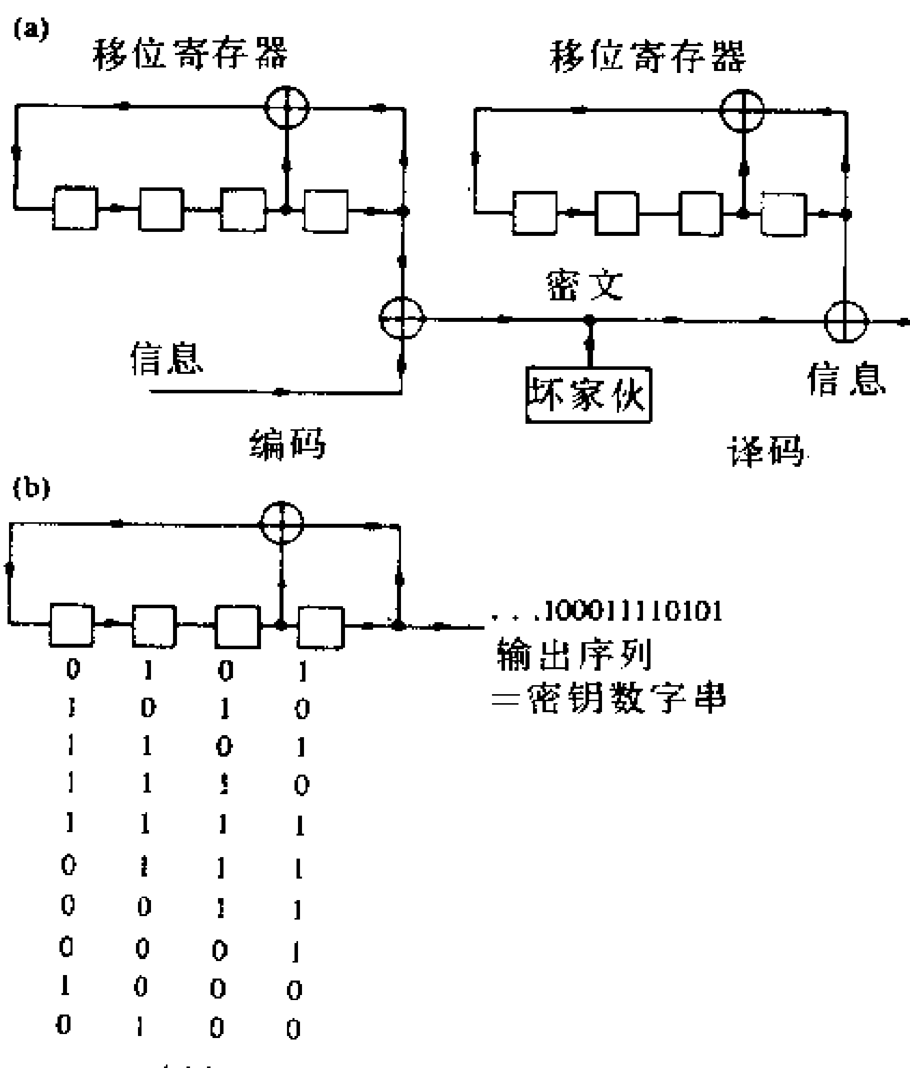


图 15

(a)一个线性反馈移位寄存器的输出可以用作密钥数字串.(b)一个四级移位寄存器的实例表明了寄存器的相继状态以及输出序列.寄存器的初始状态(0101)是一个秘密的四位密钥,它产生输出序列…111010110010001111 0101,其周期为15,再把它一位一位地加到信息上面.

(信息 X,与之对应的编码信息 Y)

——也就是若干对明文与对应的密文的组合,他现在试图找出密钥数字串 K.在线性反馈移位寄存器的情形下,搞出 K 是相当容易的.一旦知道了 X 与 Y 序列的 m 个相继匹配关系,从它们的差异中即

可决定出移位寄存器的状态，从而(只要让移位寄存器实际运转)弄明白完整的密钥(参看文献〔18〕与〔39〕). 尽管如此不堪一击，这种编码方案却依旧深孚众望，这也许是由于它的极长周期造成了人们的错觉，误认为它具有强大威力之故.

3C　非线性移位寄存器❶

可是，一旦允许在反馈线路中使用非线性元件(图 16)，情况即可彻底改观. 非线性移位寄存器系列可以做得完全合乎我们的保密要求.

有可能加以利用的一切反馈函数的总数为 2^{2^m}，当 m 数值很大时，它确实是一个异常庞大的数目. 我们希望找到一切可能的函数 f 的一个子集，它应能满足以下几项要求：

1. 能有效地把信息搅乱，从而使输出的序列类似于第 3A 节所讲的抛掷钱币所产生的序列；
2. 易于计算；
3. 易于改变(这意味着：为了改变密钥，我们只要改用一个不同的函数 f 就行).

满足以上要求的一种途径是用简单元件组建复杂函数. 此种类型的编码办法名为*乘积码*，它们很有点像传统的揉和面团的算法：

擀面团
把它对折
擀面团
把它对折
擀面团
……

❶ 译者注：中文参考书有方哲先、刘木兰、代宗铎、冯绪宁等编著的《非线性移位寄存器》，1978 年 10 月第一版.

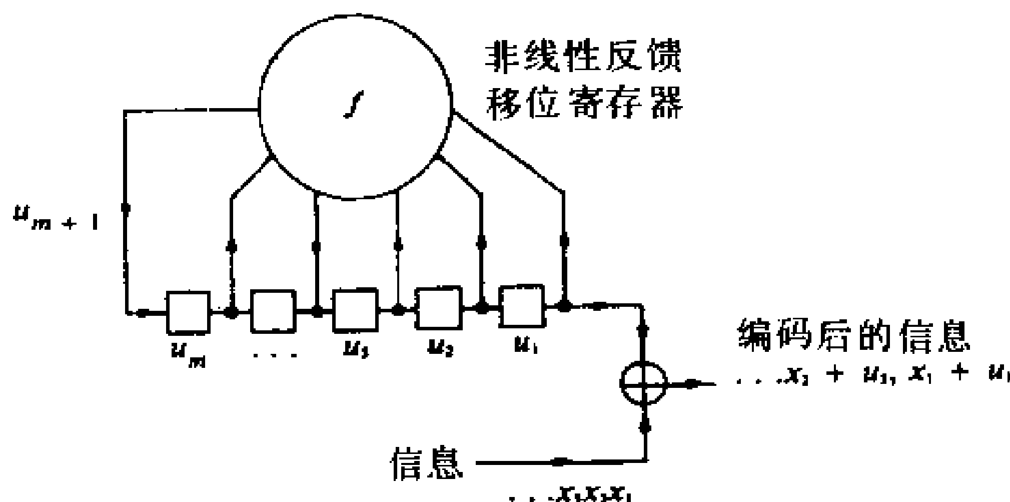


图 16

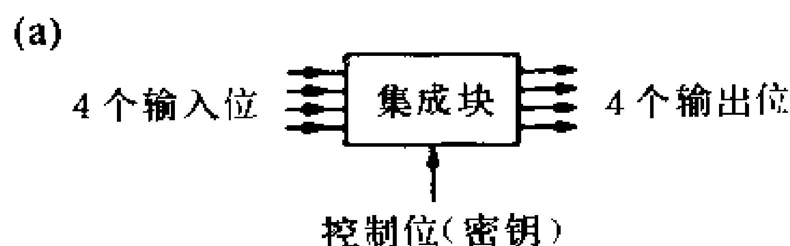
利用一台非线性移位寄存器编制密码. 这里画出来的仅仅是编码线路,不过译码线路基本上与此类似. 移位寄存器共有 m 级,开始时含有一个机密的 m 位密钥 $u_m, u_{m-1}, \cdots, u_2, u_1$. 由左端进入的 $m+1$ 位数字 u_{m+1} 是 $u_1, u_2, \cdots, u_m$ 的一个复杂的非线性函数,即 $u_{m+1}=f(u_1, u_2, \cdots, u_m)$,接下来是 $u_{m+2}=f(u_2, \cdots, u_{m+1})$,等等.

换言之,在面团上执行一系列不可交换的运算! 例如,我们可以排列以及像简单的非线性函数那样组合两类数字运算. 构筑非线性函数的一种有用元件是图 17 所示的两状态只读存储器.

为了说明这些操作在密码编制中如何结合运用,让我们考察有 32 级的移位寄存器(见图 18 底部). 基本运算是 32 位的一种排列运算(见图 18 的中部)以及图 17 的那种两状态装置,它一共有八个(都不一样),以此建成非线性函数(见图 18 的上部).

可以反复进行这些运算,例如执行 15 次,得出的 32 位数字再反馈到寄存器中(图 19),并同时用作 32 位密钥数字串,与 32 个信息位相加(图 20).

另一种运算方式是先把 32 个信息位与移位寄存器中的存储内容相加,然后再进行 15 次重新排列与非线性函数变换,最后把结果



(b)

真值表

输入	输出	
	密钥=0	密钥=1
0 0 0 0	1 0 1 0	0 1 1 1
0 0 0 1	0 0 1 1	0 1 0 0
0 0 1 0	1 0 0 1	1 1 1 1
0 0 1 1	0 0 0 0	1 0 1 0
. . . .	. . . .	. . . .
1 1 1 1	0 1 0 0	1 0 0 1

图 17

(a)带有 4 个输入位,4 个输出位,以及 1 个控制位(密钥位)的两状态只读存储器(简称 2—ROM),此种器材价格便宜,而且市场上到处有供应.(b)输出位是输入位与控制位的某种函数,其值由一张真值表予以规定.(当然也可以利用一大批更为复杂的元件,请对照《科学美国人》杂志 1979 年 7 月号第 79 页上的线路图.)

得到的 32 位作为已编码过的密文.

乘积码及其同揉和面团的譬喻最早大概是由 Shamnon〔49〕所描述.通过重新排列以及非线性线路来实现则是由 IBM 公司的一些人发展起来的,例如可参看 Feistel 的论文〔11〕,〔12〕,Feistel,Notz 与 Smith 的〔13〕,〔14〕,Girsdansky 的〔20〕,〔21〕,以及 Smith 的〔51〕. 图

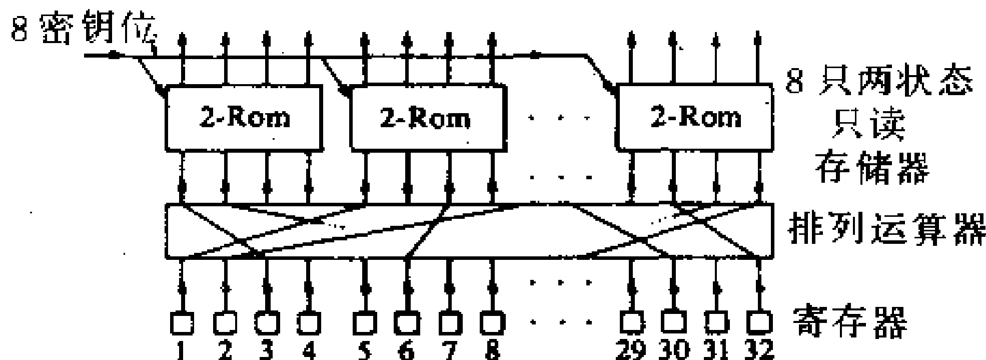


图 18

两种不可交换的运算应用于数据的 32 位. 第一种运算是一种排列,而第二种运算则是由图 17 所表示的那种非线性函数生成装置,它共有 8 只. 它们由八位密钥进行控制.

17 至图 20 的线路只是一些极为简单的例子. 读者在发明他自己的更为复杂的线路方面将不会有什么困难.

此种类型的编码装置中,受到人们密切注视的是 IBM 的罗锡弗编码方案和美国标准局颁布的数据编码标准. 罗锡弗编码方案(见文献〔11〕,〔20〕,〔51〕)利用了 128 位密钥(不是图 20 所要求的 152 位),它将数据编为 128 位一组的密文(不是图 20 中所说的 32 位一组). 数据编码标准(见文献〔3〕,〔4〕,〔40〕,〔41〕)则利用了 56 位的密钥,并把信息编为每组 64 位的密文.

不幸的是,看来对这些编码方案的保密性能研究得并不多,虽然也曾发表过一些初步的研究论文(见文献〔2〕,〔24〕). 如果密钥不长,那么坏家伙很容易通过试探各种可能密钥的简单办法来加以破译. 基于此种理由,数据编码标准本身也受到了批评,因为不相同的密钥数仅仅是 $2^{56}\approx10^{17}$(参看文献〔6〕,〔7〕,〔27〕,〔41〕,〔53〕,〔57〕). 人们迫切需要当局颁发标准以判定此种类型编码方案的保密性能.

凡是本节中所讲的所有编码方法都假定信道中不发生错误. 如果信道错误的确发生了,其后果将是灾难性的,因为好的密码都有这

样的性质：只要改变密文的一位就会在破译信息中改变全部位数的一半．为了防止这种危害，需要在编码线路的后面用上一个纠错码（见第 1 节），如图 21 所示．

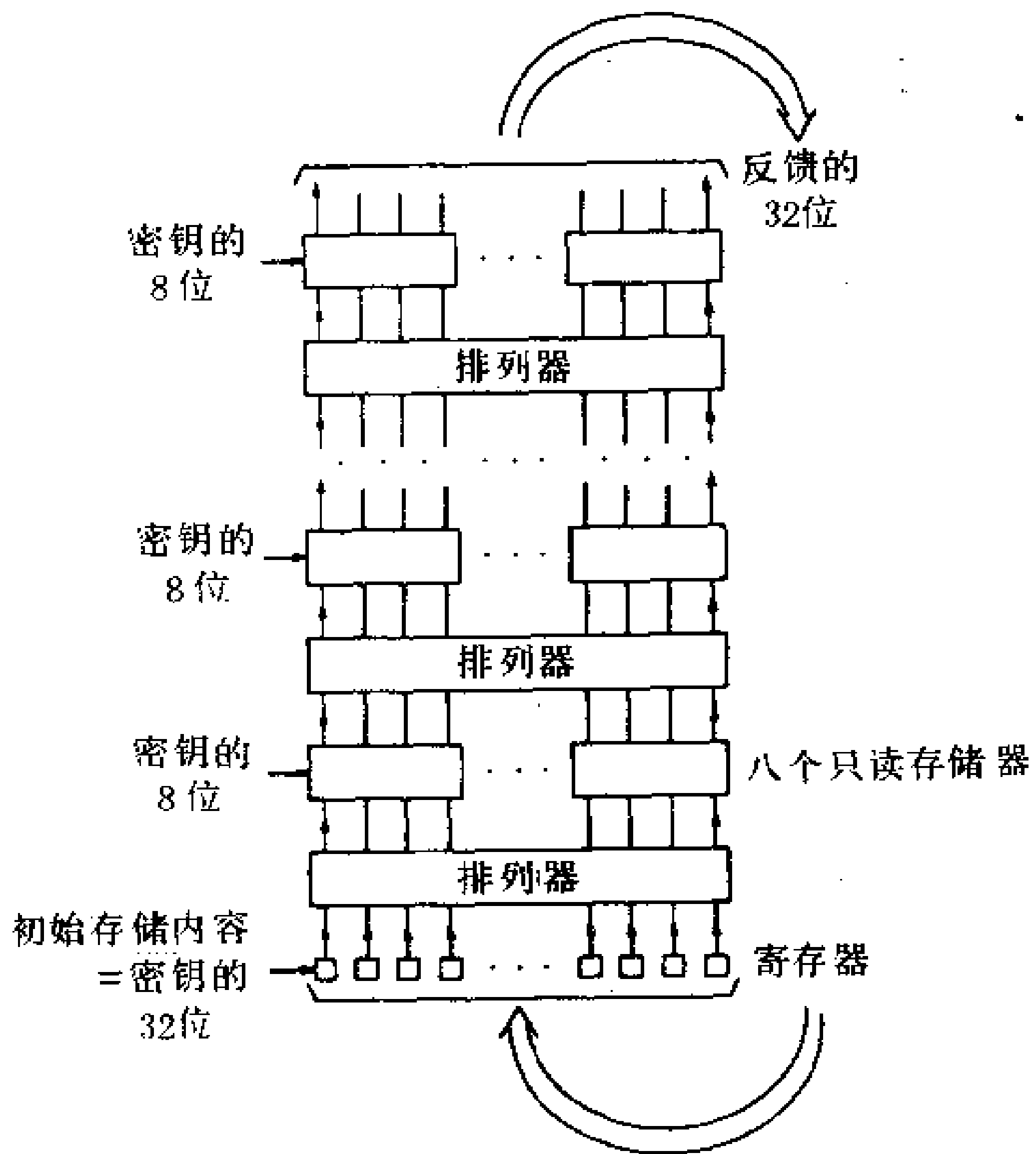


图 19

图 18 所示的运算要对移位寄存器的储存内容先后相继执行十五次，并将最后得出的 32 位反馈到移位寄存器．每个只读存储器需要一个密钥位，而寄存器的初始内存由 32 个密钥位规定，因此整个线路是由 $8\times15+32=152$ 个密钥位所规定．

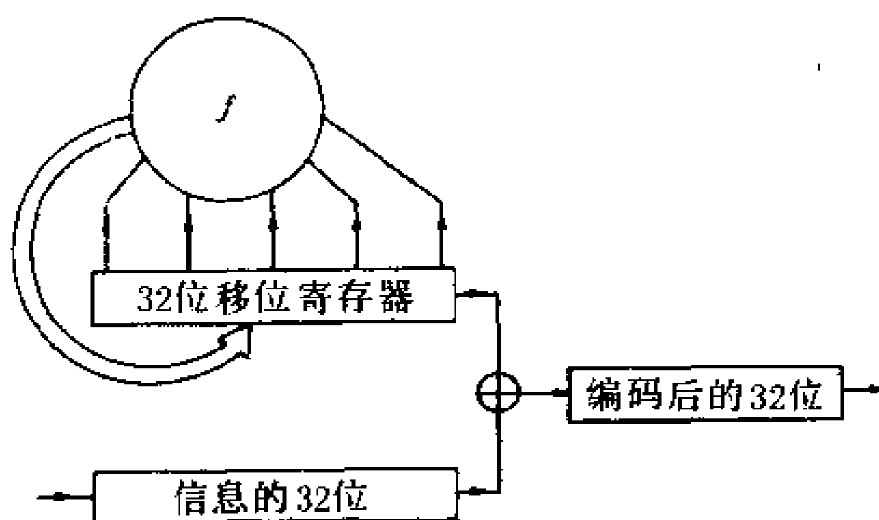


图 20

本图表明前面附图中的线路是如何编入通讯系统的. 如图 19，一个给定的 152 位密钥用来规定移位寄存器及函数 f. 需要传送的数字信息被划分为各个组，每组 32 位. 对每组数据，我们用移位寄存器的 32 位信息与之相加，这样就得出了 32 位编码信息. 然后再应用 15 次重新排列与非线性函数变换；每次都把下一个 32 位放入移位寄存器，并与数据中的下面一个 32 位相加，……如此反复进行.

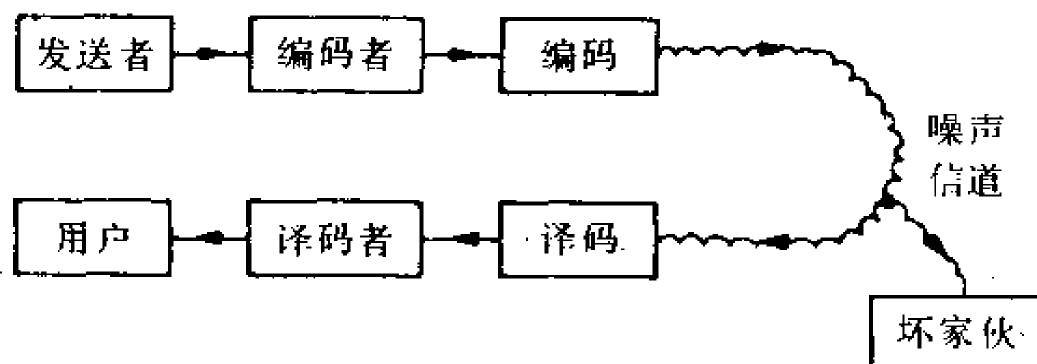


图 21

如果信道有噪声，则在编码后应当使用纠错码，因为信道中即使只有一个错误，也会使译码信息改变一半.

第 4 节
检察欺诈的密码

信道情况越来越恶化：现在它已真正被坏家伙操纵(参看图 3，第 4 部分). 他虽然应允忠实无欺地传输我们的信息，但我们不能完全信赖他. 因而我们希望信息上留下我们的签名，或者说，对信息要有一种鉴别真伪的办法，使他不可能用假货冒充真信息而不被察觉. E. N. Gilbert，F. J. MacWilliams 与本文作者在 1974 年发表的一篇论文[19]里探讨了这个问题. 有许多种情况会导致这个问题. 原先，向我们提出的人是桑迪公司的西蒙士(G. J. Simmons)，据说它同限制战略武器谈判中某些材料的生产监控有关. 但是，用赌场来说明该问题要简单方便得多.

赌场由坏家伙进行管理，他欺骗老板(是个好人)所采取的手法是少报吃角子老虎(一种自动化赌博机器)的日常收益，侵吞差额以肥己. 为了防止此事，老板提出建议，在每只机器里设置一种密钥 K，另有一种可以把每天的收益额 X 与密钥 K 都作为输入的附加装置，并由它来输出一种"签名"或"鉴定子"

$$Z = \Phi(X, K).$$

该装置把 X 与 Z 在纸带上穿孔. 然后坏家伙把纸带邮寄给老板，后者了解到 X. 从 X 与 K 重新计算 Z 值，并核查这个 Z 值是否与穿孔在纸带上的那个 Z 值相符. 如果两者一致，他就认定 X 是正确无误的.

反之，坏家伙是掌握 X，Z 与 Φ 的(这就是说，他了解装置的工作方式)但并不知道 K；他希望用另一对 X' 与 Z' 取代 X 与 Z. 如果他能通过某种方式做到这一点，即使得

$$Z' = \Phi(X', K),$$

则此种欺诈行为就不会被察觉，他就能把差额 $X'-X$ 捞进自己的腰包.

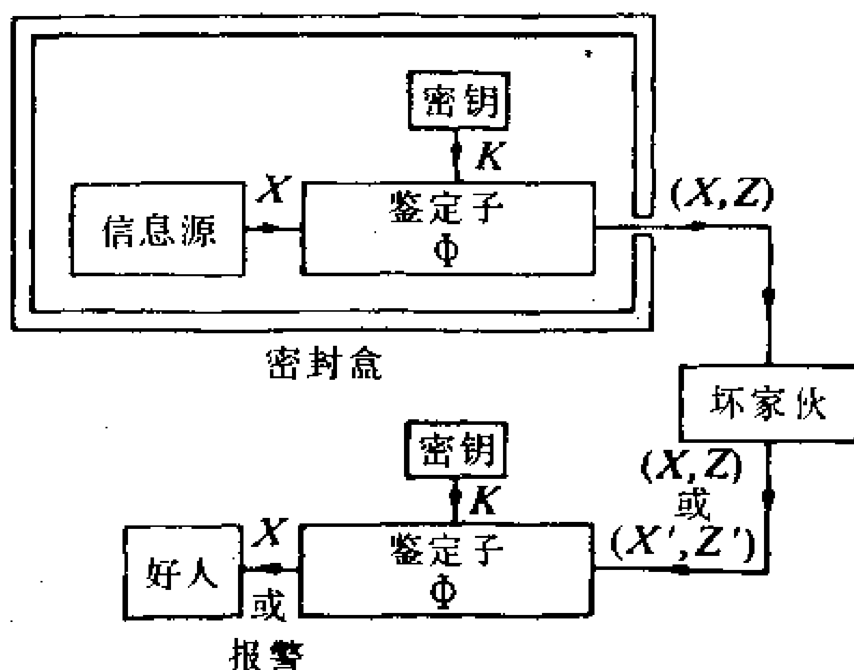


图 22

赌台老板打算在"吃角子老虎"内安装一个密封盒子.它将记录纸带上的日常收益 X,连同一个数字签名 Z,后者是 X 与密钥 K 的某一函数.于是 $Z=\Phi(X,K)$.除了 K 之外洞悉其他一切秘密的坏家伙希望用另一对数据(X',Z')来取代(X,Z).如果 Z' 能满足关系式 $Z'=\Phi(X',K)$,他将不会失风.

图 23 解释了坏家伙可能采取的分析手段.图中表明了对应于不同密钥的一切可能信息 $X_1,X_2,X_3,\cdots$ 以及鉴定子 $Z_1,Z_2,\cdots$.设坏家伙看出了信息是 X_1,鉴定子是 Z_2.通过对附图的观察,他了解到密钥必为 2,3,4,5,6 中的一个.他希望用 X_2 来取代 X_1,必须决定用哪一个鉴定子(Z_4,Z_5 或 Z_6),如果他选定 Z_5,则他幸免于被人查出的可能性最大,因其时若密钥为 2,5 或 6,他都不会被人抓住把柄,换句话说,在 5 次中有 3 次成功的机会.但如果他挑选 Z_4 或 Z_6,则他的成功机会只有五分之一.

以上论证表明,在设计一种较好的鉴定系统时,应保证使每一对信息—鉴定子都有着一大批可能的密钥.即使真的做到了这一点,也

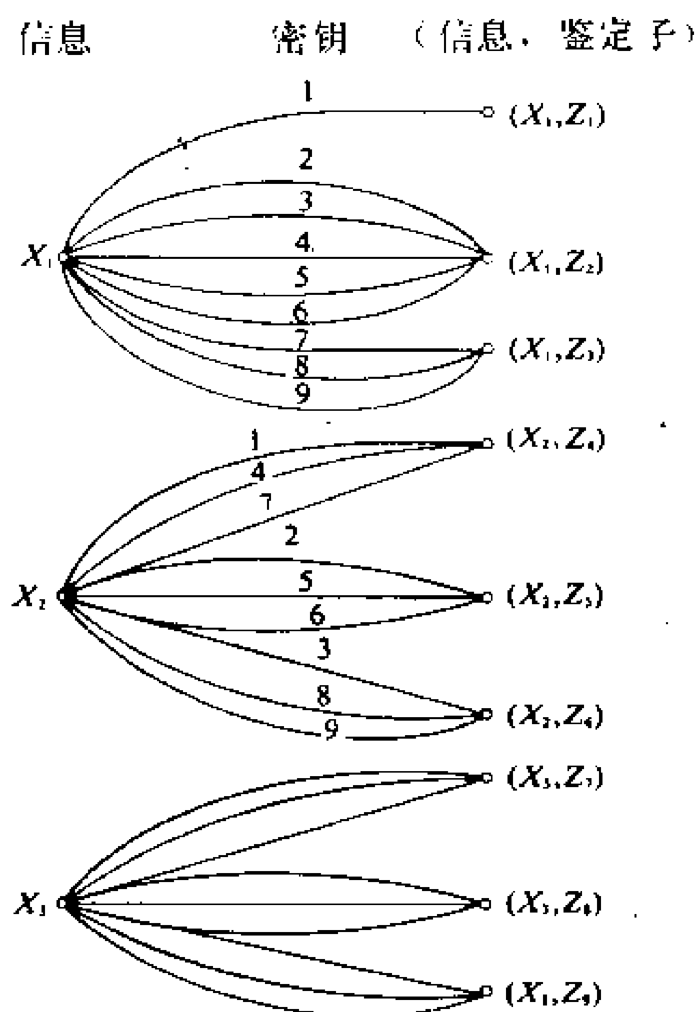


图 23

本图表明信息、密钥与鉴定子的对应关系.

很难使坏家伙的成功概率变得很小.在文献[19]中,我们证明了下列结果.

定理 2 设有 M 个可能信息 $X_1,\cdots,X_m$,N 个可能密钥 $K_1,\cdots,K_N$.如果信息与密钥是随机选取的,则在利用最优策略时,坏家伙能保证他的成功概率不低于 $1/\sqrt{N}$.

本定理的证明需要用到信息论技巧,我们将不在此处给出.如果系统设计得不好,则极有可能使坏家伙的成功概率大于 $1/\sqrt{N}$.然

而我们也能表明，应如何设计系统以使坏家伙的成功概率恰为 $1/\sqrt{N}$，根据上述定理的结论，这已是最好的可能性.（为了简单起见，我们假定 N 是一个完全平方数.）

设计需要利用第 1 节中曾定义过的，阶数 $p=\sqrt{N}$ 的射影平面（见图 24）.

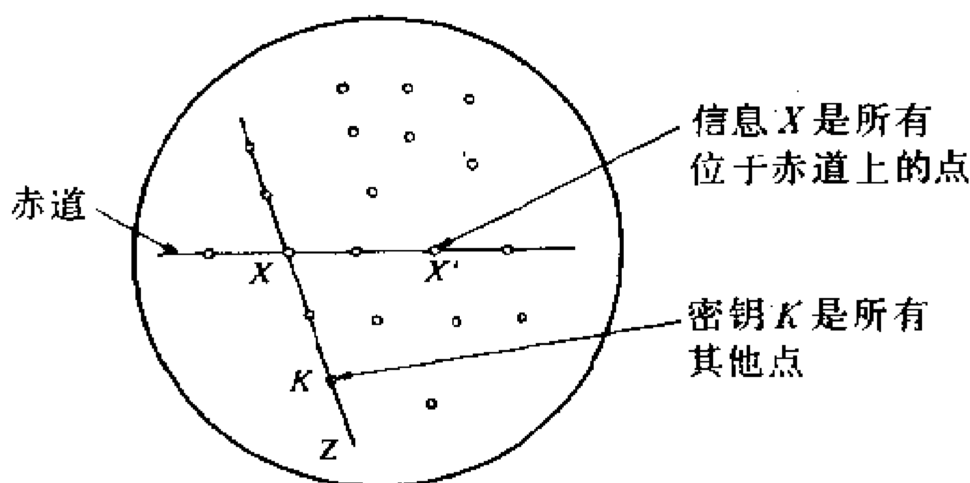


图 24

在 p 阶射影平面的基础上建立起来的最佳鉴定系统. 鉴定子 $Z=\Phi(X,K)$ 是通过 X 与 K 的直线（方程）.

请回顾一下，在射影平面上有 p^2+p+1 个点，p^2+p+1 条直线. 我们任意选一条直线，称之为赤道. 信息 X 将由赤道上的 $p+1$ 个点表示，密钥 K 则由余下的 p^2 个点表示. 最后，鉴定子 $Z=\Phi(X,K)$ 则是联结 X 与 K 的唯一直线. 该装置将把 X 的坐标与直线 Z 的方程穿孔在纸带上.

现从坏家伙的观点考虑问题. 他知道 X 与 Z，但关于密钥，他只能说，那是 Z 中余下的 p 个点之一. 如果他想用 X' 来替换 X，他所能做的事，不会比从 p 个密钥中任选其一来得更好. 由于可能的密钥（Z 上的点）与鉴定子（直线）是一一对应的关系（直线的一端要固定在点 X'），因此他的成功机会是 $\frac{1}{p}=1/\sqrt{N}$，这就是要求证明的.

上述设计主要出于理论兴趣，因为它要求拥有一大批密钥，甚至比第 3A 节提到的一次性密本还要多. 尽管如此，能够证明某种编码法是安全的（与第 3C 节以及第 5 节所叙述的编码法相反），毕竟是件好事.

值得指出，这是一种一次性密本可能毫无用处的情况. 因在那种情形（见方程 3），鉴定子将是信息与密钥的简单求和，由于坏家伙已知道信息，于是他就能立即找出密钥.

在文献〔19〕中我们也讲到一些其他的鉴定办法（基于射影空间与随机码），它们要求的密钥数量较小. 以上的分析（特别是定理 2 的证明）都假定坏家伙能毫无困难地作出他所需的一切计算. 当然实际上他要受到其计算机性能的制约. 由于这一限制，在第 3C 节中讲过的传统编码法——譬如说罗锡弗编码方案——就可能用作鉴定子. 因为它们能产生一个编码信息 Y，而后者是信息 X 与密钥 K 的复杂函数（见图 25），并且此种函数是经过特别设计的，以使得当 X,Y 与 Φ 都给出时，仍然极难找到 K. 比起我们利用射影平面的设计，这当然是一种远为实用的解决办法，可是它也有缺点，因为人们并不确切了解这种方案的保密性能究属如何.

鉴定问题的另一种解决途径是利用陷门函数，见下文第 5F 节.

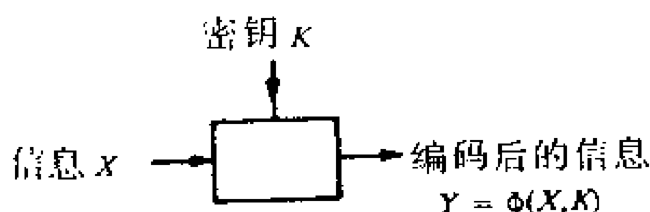


图 25

第 5 节
具有公开密钥的密码系统

5A　单行道函数

本节将描述的编码方案，其中打前锋的是简单与精巧的单行道函数. 此种函数 f 能把某些固定长度(例如有 100 位)的两进数字串映射为某些其他固定长度(例如 120 位)的两进数字串

$$f: F^{100} \to F^{120},$$

但却不知道反函数的具体求法！换句话说，如果人家告诉你

$$f(X) = 1010001\cdots11011110,$$

但是你找到 X 是没有门的，即便知道函数 f 是怎样算出的. 当然，此种说法并不十分确切，因为原则上人们可以把一切可能的 X 逐个试验，来看一看有哪个 X 的函数 $f(X)$ 可以等于已给出的数字串. 不过，这种做法实际上做不到，因为要测试的、不同的 X 实在多得不计其数. 因此我们可以说 f 是一种单行道函数，如果对任何 X 都能轻易地算出 $f(X)$，而对函数值域中的几乎一切 Y，实际上算不出 $f^{-1}(Y)$ 的话.

单行道函数肯定存在. 例如我们可以驱动一个第 3C 节提到的传统编码系统，如图 26 所示.

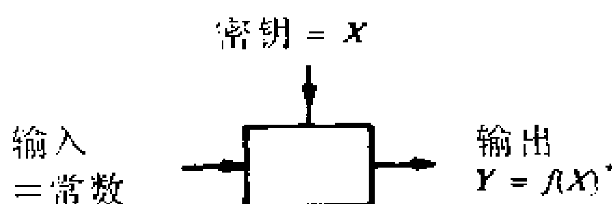


图　26

一种传统编码法(例如罗锡弗密码)可用作单行道函数. 通常的输入可令其等于一个任意常数，自变量 X 用作密钥、密码可设计得使在给定输出 Y 后，仍然无法发现 X.

这种函数对计算机有一种很出色的应用(见文献[9]，[42])．一般惯例是准许计算系统的使用者选择他们自己的口令，以便在他们

上机时鉴定用户的身份. 可是把此种口令储存在计算机文件中很危险，因为使这些文件为私人保密十分困难. 作为替代办法，人们可以利用一个单行道函数 f，并将 f(口令)值储存于文件之中. 由于 f 不能求反函数，坏家伙就无法从这文件中得出口令. 反之，有人要上机运算时，他就打入其口令 P，计算机即算出 $f(P)$，并与文件的登录事项进行对比. 这是一种非常简单，连笨人也会干的办法.（为了使它更加安全可靠，必须鼓励用户不要选择那种简短的英文名字作为口令，因为它们很容易被人猜到……）

5B 陷门函数

显然，单行道函数不能用来编制密码，因为虽然把 X 编为 $f(X)$ 非常安全，但是没有人（包括合法收信人在内）能复原 X，如同 Diffie 与 Hellman 在文献[5]中所指出. 迂回办法是利用一种他们称之为陷门函数的东西. 这是一种函数，例如

$$E: F^{100} \to F^{120},$$

它具有一个反函数

$$D: F^{120} \to F^{100},$$

并且 E 和 D 都易于计算，不过，还是无法通过分析 E 而求出反函数，除非人家告诉你反函数究竟是什么，因此 E 似乎是一种单行道函数. 于是，E 可用于编码而 D 可用于译码，这是由于

$$D(E(X)) = X$$

对 F^{100} 中的一切 X 都成立. 陷门函数还蕴含着一种思想：人们可以知道如何进行编码，但却不知道如何进行译码. 陷门函数的实例将在第 5C 节至第 5E 节中给出.

一旦我们知道怎样去发现陷门函数，我们就能建立起 Diffie 与 Hellman 所谓的*公开密钥密码系统*. 假设有一群人希望秘密交谈，于是每个人 i 选定一个陷门函数 E_i 及其反函数 D_i. 这些函数 E_1, E_2，……将全部罗列在一部公开的、类似电话号码簿那样的指南录中，但每个人都把他的反函数 D_i 秘而不宣. 当另一人 j 要向 i 发送信息 X

时，他只要简便地在公开线路上传输信息

$$Y = E_i(X)$$

就行了. 由于只有 i 才知道反函数 D_i，所以也只有 i 能够计算

$$D_i(Y) = D_i(E_i(X)) = X,$$

从而读出原有信息. 于是我们就有了一种通讯网络，可以不利用密钥而保证私人秘密不致泄露.

5C 建立在素数上的陷门函数

Rivest，Shamir 与 Adelman 是率先发现一个真正令人满意的陷门函数结构的三位学者. 这篇论文[44]已在 1978 年发表，但是马丁·加德纳在他所写的 1977 年 8 月的专栏文章[17]里已经综述了他们的方法. 由于此种方法业已引起广大公众的密切注意，这里只打算作一个简单介绍.

选定两个很大的素数（譬如说，每个素数都大约有 50 位那样长），另有一个与 $p-1$ 及 $q-1$ 都互质的数 s. 计算 $r=pq$，并找出能满足关系式

$$st \equiv 1 \quad \bmod (p-1)\cdot(q-1)$$

的数 t. 于是编码方法如下： (4)

Rivest—Shamir—Adelman 的编码方案
打印出 r 与 s.
对 p,q,t 严守机密.
编码法：
$E(x)\equiv x^s(\bmod r)=y$.
译码法：
$D(y)\equiv y^t(\bmod r)=x$.

由方程 4 及某些初等数论知识可推出 $D(y)=x^{st}=x$. E 就是(人们确信如此)一个陷门函数,因为在给出 r 与 s 之后,找到 D 的唯一已知办法是把 r 分解因子,从而找到 p,q 与 t. 然而 r 是一个 100 位左右的数字,在现阶段,对如此庞大的数字来分解因子是实质上办不到的. 所以,此种编码方案的力量在于以下事实:相对地说,找出大素数比较容易[51],而把很大的数分解因子实质上几乎不可能[25][46].

5D 建立在背包问题上的陷门函数

在 Rivest,Adleman 与 Shamir 发现其陷门函数后不久,Merkle 与 Hellman(见文献[38])发现了另一族简单函数,构成其基础的,则是所谓背包问题.

背包问题的原型是:给你一只背包以及你在一次长途徒步旅行中可能想要捎带的东西,你能否找出这些东西的一个子集合,使它们恰好装满背包? 同一类型的简单数值例子如下:

一个简单背包问题
你能否从下面列举的数字 中,挑出几个来表出 31? [10,17,9,12,40,60]? 答:当然可以,例如 31=10+9+12

由此可见 31 能表为以上这些数字中第一、第三与第四个数之和,这一事实可以记为

$$31\leftrightarrow(1,0,1,1,0,0).$$

反之，若给出(1,0,1,1,0,0)，我们只要求出第一、第三与第四个数之和，就能重新得到31.

采取此种方式，我们就可以把上述数表作为一种极为原始的编码方案的基础，如图27所示.

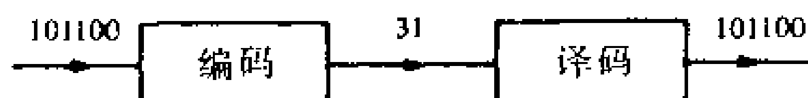


图 27

一种非常原始的编码办法，发送人与接收人都了解数表〔10,17,9,12,40,60〕.

某些数值背包问题极易解出，例如数表中的数都是2的整数幂(或者是一些迅速递增的数)时，因为它恰恰相当于找出一个数的二进制.

一个极容易的背包问题
把19表示为下述表中不同数字之和. 表〔1,2,4,8,16,32〕. 答：19=1+2+16,或 19↔(1,1,0,0,1,0).

反之，某些背包问题则解起来极其困难. 譬如说，数表中含有100个数 $a_1, a_2, \cdots, a_{100}$，每个数字都大约有40位长，而 y 则是一个42位长的数.

一个困难的背包问题
已知$[a_1, a_2, \cdots, a_{100}]$，$a_i \approx 10^{40}$.
把 $y(\approx 10^{12})$表为 a_i 之和
$y = a_3 + a_{11} + a_{12} + a_{20} + \cdots\cdots$

如果表中的数字是任意选定的，则看来没有什么现实的办法来找出其中的一个子集，使子集中的数之和等于 y，因为多达 $2^{100} \approx 10^{30}$ 个不同的子集需要考虑．如果在图 27 所示的通讯系统中利用这类数表，那么保密问题自可以令人满意，然而遗憾的是，连合法的收信人也不可能复原信息．

我们所需要的是具有陷门性质的背包问题，对合法收信者来说，解起来极其容易，但对任何其他人，都是极难解决的困难问题．Merkle 与 Hellman（见文献〔38〕）讲了几种满足上述要求的办法．下面提到的是其中某一方法的变相形式，它也曾为 Graham〔23〕与Shamir



图 28

暗中埋藏着 2 的整数幂的六个数 $a_1, a_2, \cdots, a_6$，将用于陷门背包问题．

(见〔48〕)所独立发现.

通过选取数 $a_1, a_2, \cdots, a_{100}$,我们首先形成一个容易求解的背包问题,在数的十进制表达式中埋置了 2 的乘幂(图 28 表明了一个规模较小的实例).

图 29 表明这种简单背包问题将如何用于编制密码,当然它极易被人破译,因此并不是真正我们要用的东西.

替代办法是,我们先选定两个很大的整数 r 与 s,要求存在着一个数 t,使下列关系式得到满足:

$$st \equiv 1 (\bmod r).$$

然后我们把表中的各个数 a_i 进行搅乱,其办法是,先把它们乘以 s,

信息	易解的背包	编码
1	8 3 0 1 0 5 1	8 3 0 1 0 5 1
0	2 0 2 0 6 1	
1	7 8 0 4 0 9 0	7 8 0 4 0 9 0
0	3 5 0 8 0 4 9	
1	2 4 1 6 0 1 3	2 4 1 6 0 1 3
1	3 3 3 2 0 7 8	3 3 3 2 0 7 8
		2 1 8 5 3 2 3 2

53的二进制记法为101011
这就是信息!

图 29

本图表明图 28 的容易求解的背包问题如何用于编制密码. 如图 27 一样, 只需把表中对应于信息中是 1 的各行数字相加,即可得到编码后的信息. 破译是轻而易举的: 正中间一对数的二进制记法就是信息!

然后再取模 r 的剩余数. 结果得到

$$[b_1, b_2, \cdots, b_{100}],\ b_i \equiv sa_i(\bmod\ r).$$

它们看来就很像 0 到 $r-1$ 这一段中的随机数，对不知道 r,s 与 t 的任何人来说，似乎就是一个极端困难的背包问题. 正是此种难解的背包问题才可用于实际编码. 我们把 $a_1, a_2, \cdots, a_{100}, r, s$ 与 t 秘而不宣，打印出 $b_1, b_2, \cdots, b_{100}$. 编码法如下：

陷门背包编码方案
打印 $b_1, b_2, \cdots, b_{100}$. 把两元信息 $x=(x_1, x_2, \cdots, x_{100})$ 编码为数 $E(x)=\sum_{i=1}^{100} x_i b_i = y.$ 在译码时，先形成 $ty=\sum_{i=1}^{100} x_i t b_i = \sum_{i=1}^{100} x_i a_i(\bmod\ r)$ 再解出这个易解的背包问题 以得出 $x_1, x_2, \cdots, x_{100}$.

图 30 将给出一个实例，不了解 r,s 与 t 的任何人将面临一个极端困难的背包问题，然而合法收信人却能在转瞬之间复原信息.

信息	易解的背包	难解的背包	编码
1	8301051	515862967	515862967
0	202061	679273397	
1	7804090	513438305	513438305
0	3508049	402161783	
1	2416013	427531791	427531791
1	3332078	376052641	376052641
			1832885704
	乘以 $s=324358647$ 并按模 $r=786053315$ 取余数		密码

图 30

图 28 的易解背包问题成了困难问题，为此只须乘上 $s=324358647$ 并对乘积按模 $r=786053315$ 取余数. 在破译时，我们可利用数 $t=326072163$，它是满足关系式 $s\cdot t\equiv 1 \pmod r$ 的. 编码后的信息 1832885704 乘以 t 并按模 r 取余后将得出 21853232. 把正中间的数目字 53 用二进制(见图 29)表示出来，它就是原来的信息.

5E　建立在哥帕码上的陷门函数

McEliece(见文献[37])已利用哥帕码(见第 1 节例 3)建立了一族陷门函数. 我们先确定两个数 $n=2^m$ 与 t，在域 $GF(2^m)$ 上选取一个次数为 t 的不可约多项式 $G(z)$，作出对应于能纠正 t 个差错的哥帕码的奇偶校验矩阵 H(见上文的方程 1). 由 H 计算密码的生成元矩阵，这是一个 $k\times n$ 矩阵 M，这里 $k=n-mt$. 于是，若 $x=(x_1,x_2,\cdots,x_k)$ 为信息向量，则对应的码字 $c=(c_1,c_2,\cdots,c_n)$ 将由矩阵乘积

$$c=xM(\text{按模 2 计算})$$

给出(从 H 很容易算出 M，请参看文献[35]的第一章). 关键的想法

是要搅乱 M，这可通过选取随机、可逆的 $k\times k$ 二进数矩阵 S 以及一个随机的 $n\times n$ 排列矩阵 P，并从而产生新的生成元矩阵

$$M' = SMP$$

而达到目的. 接着就打印出 M'，而对 M,S,P 加以保密. 以下即是它的编码方案：

利用哥帕码的编码方案
把 k 位信息 x 按下列变换 $y=xM'+z$ 进行编码，此处 z 是一个含有 t 个 1 的随机向量，由发送者予以选定. 译码时，要计算 $yP^{-1}=(xS)M+(zP^{-1})$， 然后按通常办法译码以得出 xS 并最后推出 x.

发信者通过改变任意选择的 t 个位上的数字以掩盖码字 xM'. 由于哥帕码有能力纠正 t 个差错，所以合法接收者能够除去这种畸变，例如可以利用在文献[35]的第 12 章中所提供的纠错过程. 可是不了解 M,S 或 P 的窃听者则只能试之又试，以破译由生成元矩阵 M' 作出的密码. 它可是一个数字非常庞大的、看来像是完全任意的线性码. 对于这样的密码，一般认为，即使不是不可能，也将是极其难以破译的.

作为一个数值例子，我们可取 $n=1024=2^{10}$，$t=50$，这时，对 $G(z)$ 就有大约 10^{149} 种可能选法，对 S 与 P 的可能选法甚至更多. 码的规模至少为 $k=1024-10\times 50=524$，此时，窃听者将面对一个字长为 1024，看来是完全任意的可校正 50 处差错的纠错码的破译问题. 欲知详情，请看文献[37].

5F 署名邮件

某些公开密码体制使发送"署名邮件"成为可能,也就是说,i 有可能通过密码形式给 j 发送一个信息,而此信息只有 i 才能发送.当密码用于传输金钱时,这种性质自然十分重要.

为了使这种想法成为可能,陷门函数 E_i 与 D_i(见第 5B 节)应当满足关系式

$$E_i(D_i(X)) = X, \text{对一切 } X \text{ 都成立}, \tag{5}$$

同样,

$$D_i(E_i(X)) = X \text{ 对一切 } X \text{ 也成立}.$$

第 5C 节讲到的素数方案显然能满足此条件.于是 i 这个人只要简单地把信息 X 编码为

$$Z = E_j(D_i(X))$$

而把它送给 j,后者即能通过以下计算

$$E_i(D_j(Z)) = E_i(D_i(X)) = X$$

而迅速地复原信息 X.只有 i 这个人才能发送此项信息,因为只有 i 才知道 D_i.因此,Z 实质上是一个 X 的署了名的版本.

满足关系式(5)的陷门函数同样也可应用于第 4 节所讲到的鉴别问题.信息 X 很易做到与鉴定子 $D(X)$结伴同来.任何当事人都知道 E,只要验证一下

$$E(D(X)) = X$$

即可证明 X 不是假冒的.然而由于 D 本身是机密的,因此坏家伙无法找到堪与他的作伪信息 X' 相匹配的鉴定子 $D(X')$.

关于署名信件的进一步资料,请参看文献[30],[44]与[45].在本书中,由 Shamir,Rivest 与 Adleman(见[47])所写的那一章谈到了这些概念对"心理扑克"的奇妙应用.

5G 结　论

其他公开密钥密码系统也有人提出来了(见〔26〕,〔33〕).在所有这些方案中都存在着一个重要的、悬而未决的问题:

它们的安全性究竟怎样?

它们像是很安全,但迄今对它们的保密力量所知甚少,隐隐然总有一种可能性存在:有一天什么人将会发明一种巧妙的破译办法.现在,已有人提出了好几种攻打素数方案与背包方案的建议,但看来尚未成为一种严重的威胁(参看〔28〕,〔43〕,〔48〕,〔50〕).

我希望通过这篇介绍性的文章,读者们将对一个激动人心与迅速发展的领域产生浓厚兴趣.我们业已看到,即使在最苛刻的条件下,仍有一些巧妙的方案可供利用,使我们得以进行机密的私人通讯.

参 考 文 献

1 Carleial, A. B. and Hellman, M. E. 1977. A note on Wyner's wiretap channel. *IEEE Trans. Info. Theory* IT—23: 387— 390.

2 Coppersmith, D. and Grossman, E. 1975. Generators for certain alternating groups with applications to cryptography. *SIAM J. Applied Math.* 29: 624—627.

3 Data Encryption Standard, Federal Information Processing Standard Publication No. 46, National Bureau of Standards, U. S. Dept. of Commerce, January 1977.

4 Davis, R. M. 1978. The Data Encryption Standard in prespective. *IEEE Communications Society Magazine*, 16(November): 5— 9.

5 Diffie, W. and Hellman, M. E. 1976. New directions in cryptography, *IEEE Trans. Info. Theory* IT—22: 644—654.

6 ________. 1976. A critique of the proposed Data Encryption Standard. *Comm. ACM* 19: 164—165.

7 ________. 1977. Exhaustive analysis of the NBS data encryption standard. *Computer* 10: (June)74—84.

8 ________. 1979. Privacy and authentication: an introduction to cryptography. *Proc. IEEE* 67: 397-427.

9 Evans, A. Jr., Kantrowitz, W., Weiss, E. 1974. A user authentication scheme not

requiring secrecy in the computer. *Comm. ACM* 17: 437—442.

10 Fak, V. 1979. Repeated use of codes which detect deception. *IEEE Trans. Info. Theory* IT—25: 233—234.

11 Feistel, H. 1970. Cryptographic coding for data-bank privacy. *Report RC*-2827, Yorktown Heights, N. Y.: IBM Watson Research Center.

12 ________. 1973. Cryptography and computer privacy. *Scientific American* 228 (May): 15—23.

13 Feistel, H., Notz, W. A. and Smith, J. L. 1971. Cryptographic techniques for machine to machine data communications. *Report RC*-3663. Yorktown Heights, N. Y.: IBM Watson Research Center.

14 ________. 1975. Some cryptographic techniques for machine-to-machine data communications. *Proc. IEEE* 63: 1545—1554.

15 Gaines, H. F. 1956. *Cryptanalysis*. New York: Dover.

16 Gallager, R. 1968. *Information Theory and Reliable Communication*. New York: Wiley.

17 Gardner, M. 1977. A new kind of cipher that would take millions of years to break, *Scientific American* 237 (August): 120—124.

18 Geffe, P. R. 1967. An open letter to communication engineers. *Proc. IEEE* 55: 2173.

19 Gilbert, E. N., MacWilliams, F. J. and Sloane, N. J. A. 1974. Codes which detect deception. *Bell Syst. Tech. J.* 53: 405—424. For a sequel to this paper see reference[10].

20 Girsdansky, M. B. 1971. Data privacy—Cryptology and the computer at IBM Research. *IBM Research Reports* 7: (No. 4), 12 pages.

21 ________. 1972. Cryptology, the computer and data privacy. *Computers and Automation* 21 (April): 12-19.

22 Golomb, S. W. ed., 1964. *Digital Communications with Space Applications*, Englewood Cliffs, N. J.: Prentice-Hall.

23 Graham, R. L. Personal communication.

24 Grossman, E. K. and Tuckerman, B. 1977. Analysis of a Feistel-like cipher weakened by having no rotating key. *Report RC*-6375. Yorktown Heights, N. Y.: IBM Watson Research Center.

25 Guy, R. K. 1975. How to factor a number. *Proc. Fifth Manitoba Conference on Numer-*

ical Math. pp. 49—89.

26 Hellman, M. E. 1978. An overview of public key cryptography. *IEEE Communications Society Magazine* 16(November): 24—32.

27 ________. 1980. A cryptanalytic time-memory tradeoff. *IEEE Trans. Info Theory.* IT-26(July).

28 Herlestam, T. 1978. Critical remarks on some public key cryptosystems. *BIT* 18: 493—496.

29 Kahn, D. 1967. *The Codebreakers.* New York: Macmillan.

30 Kohnfelder, L. M. 1978. On the signature reblocking problem in public-key cryptosystems. *Comm. ACM* 21: 179.

31 Leung-Yan Cheong, S. K. 1977. On a special class of wiretap channels. *IEEE Trans. Info. Theory.* IT-23: 625-627.

32 Leung-Yan-Cheong, S. K. and Hellman, M. E. 1978. The Gaussian wiretap channel. *IEEE Trans. Info. Theory* IT-24: 451-456.

33 Leung-Yan Cheong, S. K. and Vacon, G. V. A method for private communication over a public channel, preprint.

34 MacWilliams, F. J. and Sloane, N. J. A. 1976. Pseudo-random sequences and arrays. *Proc. IEEE* 64: 1715-1729.

35 ________. 1977. *The Theory of Error-Correcting Codes.* New York: Elsevier.

36 McEliece, R. J. 1977. *The Theory of Information and Coding.* Reading, Mass.: Addison-Wesley.

37 ________. 1978. A public-key cryptosystem based on algebraic coding theory. *Deep Space Network Progress Report* 42—44. Pasadena: Jet Propulsion Labs (January) pp. 114-116.

38 Merkle, R. C. and Hellman, M. E. 1978. Hiding information and signatures in trapdoor knapsacks. *IEEE Trans Info. Theory* IT-24: 525-530.

39 Meyer, C. H. and Tuchman, W. L. 1972. Pseudorandom codes can be cracked. *Electronic Design* 20(November 9): 74—76.

40 Morris, R. 1978. The Data Encryption Standard-retrospective and prospects. *IEEE Communications Society Magazine* 16(November): 11-14.

41 Morris, R., Sloane, N. J. A. and Wyner, A. D. 1977. Assessment of the National Bureau of Standards Proposed Federal Data Encryption Standard. *Cryptologia* 1:

281-306.

42 Purdy, G. B. 1974. A high security log-in procedure. *Communications ACM* 17: 442-445.

43 Rivest, R. L. 1978. Remarks on a proposed cryptanalytic attack on the M. I. T. public-key cryptosystem. *Cryptologia* 2: 62-65.

44 Rivest, R. L., Shamir, A. and Adelman, L. M. 1978. A method for obtaining digital signatures and public-key cryptosystems. *Comm. ACM* 21: 120-126.

45 Shamir, A. 1978. A fast signature scheme. *Report TM*-107. Laboratory for Computer Science, M. I. T.

46 ________. 1979. Factoring numbers in $O(\log n)$ arithmetic steps. *Info. Processing Letters* 8: 28-31.

47 Shamir, A., Rivest, R. L. and Adleman, L. M. Mental Poker. *Intra.*, pp. 37-43.

48 Shamir, A. and Zippel R. E. 1980. On the security of the Merkle-Hellman cryptographic scheme. *IEEE Trans. Info. Theory* IT-26 (May).

49 Shannon, C. E. 1949. Communication theory of secrecy systems. *Bell Syst. Tech. J.* 28: 656-715.

50 Simmons, G. J. and Norris, M. J. 1977. Preliminary comments on the M. I. T. public-key cryptosystem. Cryptologia 1: 406—414.

51 Smith, J. L. 1971. The design of Lucifer, a cryptographic device for data communications. *Report RC-3326*. Yorktown Heights, N. Y.: IBM Watson Research Center.

52 Solovay, R. and Strassen, V. 1977. A fast Monte-Carlo test for primality. *SIAM J. Computing* 6: 84—85 and 7 (1978): 18.

53 Sugarman, R. et al., 1979. On foiling computer crime. *IEEE Spectrum* 16 (July): 31—41.

54 Vernam, G. S. 1926. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *J. AIEE* 45: 109—115.

55 Verriest, E. and Hellman, M. E. 1979. Convolutional encoding for Wyner's wiretap channel. *IEEE Trans. Info. Theory*, IT-25: 234—237.

56 Wyner, A. D. 1975. The wire tap channel. *Bell Syst. Tech. J.* 54: 1355—1387.

57 Yuval, G. 1979. How to swindle Rabin. *Cryptologia* 3: 187—189.